A general methodology for designing and developing Intelligent Database Decision Aids, with application to medicine

SUSAN AMANI JACKSON

A thesis submitted in partial fulfilment
of the requirements of De Montfort University
for the degree of Doctor of Philosophy

March, 1995

De Montfort University in collaboration with
Leicester Royal Infirmary

ABSTRACT

Susan Amani Jackson
De Montfort University in collaboration with
Leicester Royal Infirmary

A thesis submitted in partial fulfilment of the requirements of
De Montfort University for the degree of Doctor of Philosophy
March, 1995

A general methodology for designing and developing Intelligent Database Decision Aids, with application to medicine

After more than twenty years of development effort in expert and knowledge-based applications, there are indications of a growing uncertainty in the practical potential of such systems, especially within medical domains. Many difficulties have arisen from attempting to model human experts. These are particularly evident when considering unformalised or non-standardised domains which are common characteristics of many specialist medical fields. Moreover, little attention has been paid to the prospective users of these systems, the tasks which are routinely undertaken or the environment in which the users must operate. These factors have all contributed to the continued lack of success of such systems.

This research reviews the difficulties encountered during the development of knowledge-based systems and conventional systems. From these studies, the importance of fully considering end-users and their needs became evident. It also became apparent that currently, there is a lack of techniques available to medical investigators which would allow them to quickly, easily and thoroughly analyse the information they collect during their research studies. However, the ability to undertake such reviews is crucial if consultants are not only to extend their knowledge of their domain through exploration but if they are also to evolve agreed operational practices. This standardisation of approach would lead to a rationalisation of the tests and procedures routinely undertaken, which in turn would result in the saving of time, money and patient discomfort.

Consequently, this research also examines the intended user group, the typical procedures followed and the common tasks undertaken during clinical trials, as well as the environment in which the user group operates. These studies uncovered the typical facilities and assistance required by such investigators. From this information, a general methodology, characterising the processes involved in the construction of a generic Intelligent Database Decision Aid (IDDA), was developed. A suite of computer-based tools then evolved to facilitate the tailoring of such a system by a domain expert, who may be a naive computer user, for a specific investigation. These tools would thus give total control of a study to the domain expert and permit an IDDA system to be quickly and easily constructed for each new investigation.

The approach was evaluated by utilising test cases drawn primarily from the medical domain. However, as the methodology was based upon commonly accepted investigative procedures, it was also reviewed in other domains to test its wider applicability. All of the IDDA systems were successfully constructed and the feedback obtained from the trials was very positive, both for the approach adopted and the various IDDA end-systems produced. Therefore, the general methodology proposed by this research has been shown to be effective and its benefits can now begin to be realised.

Acknowledgements

I would like to offer my sincere thanks to all those who made this thesis possible.

First and foremost, I am indebted to Professor Peter Thewlis. His patience, guidance and understanding not only as a supervisor but also as a friend, greatly influenced the present work. Secondly, I would like to thank Professor Derek Teather for his willingness to offer regular comments and advice. Similarly I would like to express my appreciation of the clinical expertise and assistance provided by Mr Harding and the Orthopaedic Department, Leicester Royal Infirmary Hospital.

I also wish to express my gratitude to Liz Grayson, Trevor Townsend and Brian Kiely for their friendship, wit and ready willingness to read and discuss the various sections of this work.

Last, but not least, I am especially grateful to my parents and my brother for their continued encouragement, good humour and support throughout the development of this thesis.

Contents

Figures and Tables

Table of Abbreviations

CAT	Computerised Axial Tomography
CBF	Carbon fibre
GUI	Graphical User Interface
HCI	Human Computer Interaction
IDDA	Intelligent Database Decision Aid
ISP	Ill-structured problem
LRI	Leicester Royal Infirmary
MOT	Ministry of Transport
WSP	Well-structured problem

Chapter 1

Overview of the Problem Domain

1.1 Introduction

Over the past century, the quantity of available information has grown dramatically, creating a major difficulty as the innate human capability for information storage, retrieval and analysis has remained roughly constant (Parsaye and Chignell, 1993). Consequently, there has been a continual desire to find better methods for managing and manipulating the mass of information in a more efficient, productive and beneficial manner.

The evolution of computer systems, both in terms of hardware and software, over a similar time period and the impressive achievements of this technology in certain areas, has led some people to believe that the solution to their problems of limited human capabilities resides within the scope of computer technology. New research fields have emerged, for example artificial intelligence, expert systems, and knowledge-based systems. Gradually, these have combined with older disciplines such as psychology, in an attempt to discover how humans process information, the limitations of the human approach and what leads to and what constitutes intelligent or expert behaviour.

One of the realisations from this research is just how impressive the human mind really is. Certain tasks and abilities which are taken for granted, for example, vision and language understanding, have proved very difficult to computerise. Consequently some of the initial enthusiasm and wild predictions have been tempered with a growing recognition that 'people and computers both have vast but, for the foreseeable future at least, somewhat different potentials' (Parsaye and Chignell, 1993).

The research described in this thesis reviews these different potentials and reflects how the strengths of each could be combined to enable more informed and, hence, better decisions to be reached than either a human or a computer could achieve on their own. Firstly, there is a discussion of those difficulties that have beset expert system development and how they have been further compounded when such systems were implemented within medical fields. The evolution of, and the issues raised during, conventional systems development are also reviewed as are the current practices involved in typical clinical investigations. From these studies, the structure and component parts of a generic IDDA system for clinicians to explore their field became apparent. A methodology then emerged which, when computerised, enabled naive computer users from the information they provide to develop and tailor an IDDA system for their own specific investigation. Although this methodology has primarily been evaluated by using examples from the medical domain, it is based upon commonly accepted investigative procedures and thus it is anticipated that it will prove to be equally applicable to investigations undertaken in other fields.

Therefore the aim of this research is to characterise the processes involved in the design and construction of intelligent database decision aids (IDDA) and to develop tools to facilitate the construction of such aids by domain experts who are naive computer users. The following sections provide a brief overview of the various issues and concerns inherent in these tasks. They are then discussed in more detail in the subsequent chapters of this thesis.

1.2 Problems of expert system development

Traditional Artificial Intelligence (AI) concentrates on search strategies and explores the mechanisation of thought through various games, with chess being perceived as the supreme challenge. The reason for focusing on such domains was that 'they require little contact with the outside world' (Brooks, 1991) and therefore could be considered in a closed environment divorced from any external influences. The implementations of intelligence that evolved, however, bore little resemblance to the mechanisms used when humans are confronted by similar problems. In addition, with speed as the critical criteria, the most successful programs of the day tended to rely on technological advances rather than a closer imitation of human reasoning processes (Brooks, 1991; Davis, 1989b). Consequently, although successes were recorded in games such as chess, others with much larger search spaces such as 'Go' continued to be too problematic for the techniques currently available. Thus the advances made, although important within the AI field, were not particularly relevant with respect to practical applications operating in the 'real-world'.

With expert systems however, there was an attempt to implement in practical systems some of the theories evolving from AI research. Boose (1985) defined an expert system as 'a computer system that uses the experiences of one or more experts in some problem solving domain, and applies their problem-solving expertise to make useful inferences for the user of the system.' The knowledge upon which these expert systems is built is generally gathered from the human experts in the form of rules of thumb or heuristics. Heuristics enable human experts to draw together their knowledge to make educated decisions about problems they have not previously met. The common aim of expert systems has been to simulate those elements of human reasoning in which human experts are highly competent, then to apply these systems, in a decision support role, to aid lesser experts in solving the same problems.

However, expert systems have not been as successful in their general application as their early successes may have promised. A number of different reasons have been proposed in an attempt to explain this failure. These difficulties can generally be allocated to one of three major factors: the people involved, the development method adopted, and the monetary costs. The following list is neither exhaustive nor indicative of the priority, it merely represents a sample which can demonstrate the breadth of the problems facing developers of expert systems.

People:

a) the lack of available and willing expert(s) - experts may wish to guard the exclusiveness of their expertise or may be concerned over the possibility of such a system making their hard earned skills redundant (Irgon et al, 1990)

b) the inability of experts to articulate their knowledge or the attempt to automate tasks which require common-sense - currently no expert system adequately deals with common-sense (Keyes, 1989)

c) the involvement of too many experts - resulting in an inability to achieve a consensus or a consistently high level of expertise

d) the lack of management support - the view that is often held is 'why change from tried and trusted methods'

e) the lack of user support - user acceptance is a crucial factor in the success of any computer system, they must in fact become advocates for the system (Duchessi and O'Keefe, 1992). 'Users must first perceive a need. Then they must see the benefit because we can't force them to use the system' (Keyes, 1989)

f) the lack of appropriately trained and able knowledge engineer(s)

g) the lack of user confidence and trust in the system - the output from an expert system can be difficult to understand and the idea that systems based on rules can be made more intelligible by literally presenting the rules used in making a recommendation is now largely discounted, 'explanation will commonly entail an understanding of the deeper justification for the rules as well' (Southwick, 1991)

Development:

a) the lack of users in the development team

b) the inability of the implemented expert system to fulfil operational requirements in the work place

c) the lack of adherence to a development life cycle methodology - thus the actual requirements of the system, the progression and the achievements are hard to ascertain. In addition, the evaluation and testing stages within 'real-life' environments have tended to be overlooked (de Dombal, 1984)

d) the selection of an inappropriate problem - the 'real-world' knowledge which is utilised during problem-solving must be capable of being formalised into something tangible, 'we operate by reducing any domain of concern to a collection of elements that can be related in reasonably definable ways - among which you can make clear distinctions' (Davis, 1989a)

e) the setting of unrealistic goals - the problem area selected was too ambitious in scope (Davis, 1989a)

f) the inability of the expert system to handle problems which were not explicitly recognised in their development - expert systems are widely criticised for being too 'brittle' and practical experience has demonstrated that they are rarely more than 20% of a complete solution to 'real-world' problems (Fox and Krause, 1992; Hayes-Roth and Jacobstein, 1994). Consequently, a situation could arise for which the appropriate specific knowledge had not been encoded, resulting in the system suddenly collapsing. As Fox and Krause (1992) point out, this is particularly worrying if the collapse occurs without any obvious sign being apparent to the user

Costs:

a) the lack of any pay-back - there was no obvious benefit from using an expert system (Keyes, 1989)

b) the monetary, resource and time commitment required to design and implement expert systems - it is not unusual for a large system to absorb 10-25 man-years of effort (Hart, 1982)

c) the continuing cost of maintaining the knowledge-base - XCON, often viewed as a successful expert system, costs DEC upwards of $2 million a year to keep up to date, 'even this injection of money can't keep the system current with the proliferation of engineering changes that XCON must keep track of' (Keyes, 1989).

Many of the above issues and their relationship to this research are discussed in more detail in subsequent chapters. For example, Chapter 2 considers the problems of knowledge acquisition and the people involved in that task whilst Chapter 3 reviews the development process and the importance of factors such as end-user involvement and user acceptance of an implemented system.

In recognition of the above difficulties, there has now been a suggestion that systems should support those aspects of reasoning in which most humans are weak as a result of their inbuilt cognitive limitations rather than giving assistance in areas where human experts are "good" (Woods ,1986; Miller, 1986; Kidd ,1987). This had emerged from the realisation that most operationally successful systems acted in an ancillary capacity by making predictions from the analysis of observed data rather than relying on 'guesstimates' of experts, which had been shown to be highly unreliable in practice (de Dombal, 1984; Fieschi, 1990).

In addition, it has been argued that to utilise a computer system to best advantage, the users must understand the actions of the system, especially with respect to the domain (Davis, 1989a). This would imply that for expert systems, which are naturally complex due to their composition, the operators who are in fact best equipped to use such systems are the actual domain experts themselves. This is contrary to the commonly held view that novices are the major beneficiaries of this technology.

However, it should be stated that domain experts do also experience the limitations of human reasoning. The Newell and Simon (1972) explanation of 'limited rationality' is one of the basic principles which has led to an understanding of clinical reasoning. This belief is that the human capacity for reasoning is limited by an ability to only cope with a restricted number of facts at a time. Consequently, it is not possible for anyone, including an expert, to work efficiently with all the available knowledge on a subject, or with all the facts which can be gathered.

Even if experts could utilise all of the available knowledge, there is a drawback. For although knowledge is generally assumed to be beneficial in decision making, acquiring complete knowledge is rarely a practicable proposition, especially when decisions involve considerations of the future. In the unique situations were it is possible to remove uncertainty, it may be just too expensive or harmful to find out the complete truth. Consequently, there is a need to represent simply and analyse quickly a selection of those facts which, to a user, appear to be the most important. This will help remove part of the uncertainty by providing some additional relevant information about the problem. Therefore, although the true situation cannot be determined precisely, the users should be more knowledgeable than they were without the information (Moore and Thomas, 1976; Fieschi, 1990).

This ability would be particularly beneficial in medical fields where information has expanded to the extent that the assimilation of all the published data is felt to be impossible. Physicians and other health professionals often find that they either do not have sufficient access to necessary information, or do not have the time to adequately acquire, review, and analyse the appropriate information. Consequently, they are working with much greater uncertainty, which is resulting in both poorer quality of care and increased health costs (Hayes-Roth and Jacobstein, 1994). Yet it is not only the provision of appropriate systems to support medical personnel that is lacking, there is also a need to include information processing modules within medical training courses. Shumway et al (1990) agree, 'unfortunately, the science of information processing, which should be a fundamental part of medicine and medical education is currently not integrated into the majority of medical curricula'.

It is evident therefore that medical fields could benefit from the use of appropriate computer systems. However, the nature of the job, the depth and breadth of the subject area, the working environment and the culture that pervades old disciplines such as medicine, have created additional difficulties which did not receive enough attention during the implementation of the initial systems. Thus the final outcome of all these problems has been the lack of successfully developed and implemented medical expert systems. This has resulted in questions being asked as to the suitability and feasibility of such systems in the medical environment. As Lipscombe (1989) states, with 'intelligent knowledge-based systems or expert systems, there are indications of a growing uncertainty regarding the practical potential of these programs. The reason for this uncertainty, at least in AIM [Artificial Intelligence in Medicine], is an embarrassing lack of success in the development of usable systems. After something like 20 years of development effort, for example, there are no significant medical programs in use, outside of a research environment'. Some of these difficulties are briefly described in the following section.

1.3 Problems inherent in medical computer system development

A number of the factors responsible for the failure of expert systems to fulfil initial expectations have already been highlighted. However, within the field of medical expert system developement, there are in fact a further two fundamental reasons given for failure:

a) that medical diagnosis and therapy are not sufficiently defined with explicit rules to be represented in an expert system (Colleste, 1992; Johnson, 1983),

b) the wide scope encompassed by the subject material and the inherent uncertainties in medical information (Kassirer and Gorry, 1978).

Thus, for medical expert system development, key elements were missing: the ability to acquire and to represent explicitly all of the knowledge required.

Moreover, there was no evident benefit, neither in terms of professional advantage or financial incentive, for physicians to learn to use such technology, nor was there money available to finance the development and, more importantly, the maintenance, of these systems. 'Many practitioners perceive the rewards of using clinical systems as currently available are simply not worth the extra effort of using them. Data entry (and sometimes retrieval) is much more cumbersome than using a pen or gabbling into a dictating machine' (Vincent, 1993).

Rector, (1989) summarises four further factors that affect the development of medical computer systems:

- doctor's primary attention should always be focused on patients rather than computer systems,
- doctors work in brief consultations under severe time pressure. In most medical fields average consultation time with a patient is only five to fifteen minutes,
- the wide diversity that exists in the users' clinical skills. Medical practice includes many transient doctors - housemen and registrars in training, locums filling in, etc. Each doctor has an idiosyncratic range of clinical skills with hard won expertise in some areas and serious gaps in others,
- the wide range of experience with the system which is to be expected. If a system is successful at becoming a routine part of practice, many doctors will be using it four to six hours per day, four and a half days per week. On the other hand, the system must also be accessible to users who meet it for the first time.

Yet with all these difficulties, there is much evidence that physicians do require assistance to make the best decision when contemplating not only diagnoses but also treatment strategies. However, this evidence is often scattered and its potential is often overlooked. Thus, as Christopher Hart (ophthalmologist consultant at Bristol Eye Hospital) states, 'the big problem in modern medicine is too many people making assumptions on insufficient data. Essentially medicine remains an art. I do tests and make judgements on what I see' (Fisher, 1994).

Konner (1993) compiled a number of such cases in his book 'The Trouble with Medicine'. One example being the value of a extracranial-intracranial surgical procedure, as opposed to a drug therapy programme, to alleviate a clogged artery high in the throat. Reports were published, describing the success of the surgical procedure resulting in more and more patients being treated by this technique. However, no-one had really proved that it worked. Eventually, Dr Barnett undertook an international study comprising of 1400 patients randomly assigned to be either treated by the drugs or by surgery (in Konner, 1993) . The outcome was difficult for many doctors to believe: there had been no advantage in using surgery. Dr Barnett stated, 'this stunning reversal of what everybody thought made a lot of people think that a lot of surgical procedures had to be evaluated with controlled studies. I think it was in that way, for those of us involved with the nervous system, a bit of a landmark study. It taught us that we simply couldn't go on believing that what we were doing was right without proving it' (Konner, 1993).

This is only one example but it demonstrates the point that treatment strategies can be adopted widely without being properly tested first. This is possible because, 'there is no procedure for formally scrutinising operations or diagnostic tests, unless they involve new substances' (Konner, 1993). The adoption or rejection of a new technique is left to the judgement of experienced physicians. This results in instances where the treatment given is actually ineffective or harmful. Moreover, if an appropriate analysis of the technique were to be carried out before its full scale introduction, it is likely that the outcome could also be predicted on the basis of a patient's status and record.

There are other factors which influence a physician's decision. One is the threat of being sued for malpractice, which is particularly prevalent in the U.S.A., where physicians are compared with the 'standard practice' of the day. The other is the natural group pressure to conform and adopt fashions which seem appealing and for which rational justifications can be made. Hence, the result is the widespread adoption of a new procedure without clear evidence that it works. 'Lobotomies came and went even more dramatically but in essentially the same way as the tonsillectomy fad during those decades and the hysterectomy fad a little later: credulous doctors sent their patients to over-confident surgeons, whose standards of practice were shaped by colleagues' anecdotes and by their own experience rather than by rigorous scientific research' (Konner, 1993).

This practice of not insisting on controlled trials being undertaken also effects the training of medical practitioners since the procedures they are taught and will continue to use, have never been properly tested. Dr Eddy discovered this fact during his medical training and was quite disturbed by it as he explained to Konner (1993): 'In fact very little of medicine has been carefully evaluated in well designed, well controlled studies. It's really quite amazing, but after hundreds of years, I would estimate that only about ten to twenty per cent of medical practices have been evaluated properly. What that means for the patient - and not just the patient but for the physician - is that for a large population of practices we really don't know what the outcome or what the effects are.'

Furthermore, if more operations are carried out than is necessary, not only is the risk to patients increased but also the costs of health care rises. Rising costs will also occur if more expensive procedures are selected when they result in little or no extra benefit being obtained. For example, magnetic resonance scans replaced CAT (computerised axial tomography) scans as the most common technique to use. However, although they were better for some tasks, they were not three times better even though they cost three times more (Konner, 1993).

Therefore it seems that one of the more urgent areas in which physicians can be given assistance by computer technology is in the undertaking and running of randomised clinical trials. The benefits could be seen in terms of improved medical practice, reduced health care costs, minimised risk and discomfort to patients, and a clearer understanding of the medical decision process.

The benefits of a more explicit understanding of this decision process are undeniable. As Kassirer and Gorry (1978) pointed out, 'an accurate delineation of principles and strategies should improve the process of developing clinical reasoning in young physicians. Such principles and strategies could also be used in the development of better measures of the quality of medical decision-making and could, therefore, improve medical care.'

The changes imposed on the working practices of the medical unit by the introduction of such a system could be diminished dramatically by these systems adopting currently used manual procedures for data capture. Generally medical information is held in the form of patient records. Within specialist fields, or indeed clinical trials, where the medical interest is more focused, these records could be standardised for the study and thereby enable statistical analyses to be undertaken on the gathered information. This technique of using statistical methods to attempt to extract rules implicitly from the collected data is different from the AI techniques which attempt instead to explain the rules (Fieschi, 1990).

Medical records have already been shown to be useful in many aspects of clinical work. As Safran et al (1991) observed, they can increase access to information, improve physician behaviour, improve the quality of patient care and provide an opportunity for outcome-based research. Moreover, a number of the drawbacks of manual records (for example illegibility, inaccuracy, incompleteness and being disorganised) could be removed by their automation.

Computerisation would also allow flexible access to the information and this could include facilities for investigating possible relationships between incidents and variables. Also, as Parsaye and Chignell (1993) explain, the value of information rises dramatically with the size of the information store (or database). This is due to the increase in the number of connections and relationships between objects that can be identified when using a larger information base. Blascovich (1987) agrees stating, that 'only the creativity of the user will limit the informational value of the database management system once the record items have been selected.' Therefore it is important to provide appropriate techniques for selecting and analysing the data to enable the investigator to make sense of and full use of the information that has been gathered (Piatetsky-Sharpiro, 1994).

The results of any analyses can then be circulated for further comment and review. The value of giving feedback, in the medical field in particular, has already been demonstrated. Wennberg and Gittelsohn (1969) undertook a study of the number of tonsillectomies carried out in Vermont, U.S.A., and discovered a 13 fold difference between the highest- and lowest-rate. They then informed doctors how they stood in relation to other doctors. The outcome of this feedback was a decline in the rate of tonsillectomies in Vermont at a rate far exceeding the national decline. After careful analysis, Dr Wennberg 'attributed much of this [decline in popularity of tonsillectomies] to the review of the procedures done and feedback of information to practising doctors' (Konner, 1993). De Dombal (1984) and de Dombal et al (1991) have also studied the benefit of giving performance feedback to physicians, although in these cases it involved the procedure for diagnosing appendicitis.

Although some people are sceptical about physicians accepting such feedback, the indications from those studies that have been undertaken have been very promising and doctors have actually welcomed them. Dr Hart in fact believes that 'good doctors are those who are prepared to measure or let others measure, how bad they are; or, more constructively, are prepared to accept that their work can be convincingly improved only if they are prepared to start by measuring its outcomes, errors and omissions' (Konner, 1993).

Therefore the only real factor that is preventing the realisations of all of the above benefits is the lack of quick and easy methods for carrying out the required investigations. The capabilities of the computer seem to make it ideally suited to providing the necessary assistance, if appropriate software could be developed. This software must accommodate all the inherent difficulties of a medical environment as well as all the facilities that the medical personnel may require to manage and analyse the assembled information.

The objective should be to support the physicians in their tasks and not to insist that users should be directed, or dictated to, by computers. Therefore, 'placing more emphasis on the concept of a man-machine team will lead us to concentrate more on developing machine expertise in those tasks where man needs them more. In a sense, we are proposing to shift emphasis to what computers should do rather than what computers can do' (Ben-Bassat, 1985). This proposal of teamwork is in contrast to the traditional AI strategies whose aim is to replicate and replace human ability (Parsaye and Chignell, 1993). However it is believed that this approach will lead to a result that is better than if the individual members of the team, i.e. user or machine, had worked in isolation. Nevertheless, to achieve this goal, there needs to be an identification of tasks followed by the correct allocation to each member of those tasks in which one has a relative advantage over the other.

The next section briefly describes the type of computer system that would be relevant for specialist medical fields undertaking clinical trials.

1.4 Requirements of a medical computer system for a specialist field

By basing the computer system on data that is already collected in a specialist field, the necessity to acquire and rely on the detailed knowledge and heuristics gained from the expert is removed. Knowledge acquisition and representation are two of the major problems in developing expert systems. Furthermore, as expertise, in any field, is based on increasing levels of 'automaticity' of physical and mental processes, the more accomplished an expert is the harder knowledge is to obtain. A number of these difficulties are described in more detail in Chapter 2.

The problems of describing and explaining one's actions and the reasons behind them are further compounded when the experts themselves are uncertain of the outcomes of their own individual actions. If you do not know precisely why a particular event occurred, how can you be sure of reproducing the same event?

Within many specialist medical fields, doctors are unsure of the outcome of the treatment they prescribe. Decisions over the best treatment for a specific patient are therefore exceptionally difficult. One of the major problems is that many treatments do not show their "success" or "failure" until a number of years after the initial treatment. Furthermore, people can react in different ways to the same treatment and other outside influences, i.e. their working environment and life styles, all play a part in determining the final outcome. The result is that many doctors use their own preferences or the fashionable treatment at that time. However, uncertainty in a domain can be reduced considerably if a decision-maker can inspect the results of similar situations that have occurred in the past.

What is evident is that there is a mass of valuable information being gathered within specialist medical fields that currently cannot be (and is not being) exploited to its full potential due to the restrictions of analysing it by hand. This is a very laborious, mundane and time-consuming task and with the prevailing working conditions, i.e. frequent interruptions, the limited capacity of the human mind to concentrate intensely for long periods of time and the complexity of most analysis techniques, errors and omissions occur. Hence inaccurate and/or misleading results are produced. Consequently, only a small fraction of the possible benefits from the information is actually being realised (Parsaye et al, 1989; Parsaye and Chignell, 1993). Thus there is a need to provide appropriate storage, retrieval and analysis mechanisms for physicians who find themselves in this situation. These tasks are in fact ideally suited to computer technology, which never tires, loses concentration or makes mistakes during long and complex calculations. As Yntenna and Torgenson (1961) propose, a possible 'solution is to have the machine assemble and present to the man the facts he will probably want in reaching a decision, allow him to call for additional information if he thinks he needs it and require him to indicate to the computer the action on which he decides [to ensure that the information is kept up to date and that future investigations are also accurate]'.

Consequently, the emphasis must be towards developing computer systems that can aid the users in decision-making and not producing systems that attempt to reach the decision for users. 'A human-centred system provides an environment in which as much decision-making is given to the user as possible; in which increase performance is a function of the growth of the user's skills in manipulating the system; in which an increase level of human skills is thereby fostered; and in which personal responsibility has real and direct applicability in achieving and maintaining the productive process' (Young, 1989).

Provided that the techniques offered are appropriate for the tasks being undertaken in the work environment and that they are easier to use for the same output as existing systems or only slightly more difficult to use for enhanced functionality, physicians have expressed a willingness to enter the required data (Vincent, 1993; Safran et al, 1991). This is especially relevant now, as more pressure grows from quality reviewers and risk managers for clinicians to improve their recording of data (Cushing, 1991). This has also led to a need for standardising the information gathered to enable valid comparisons to be made. As Cuesters et al (1992) state, 'it has to be stressed that the medical record is the single most important tool within a patient care system.' This is especially true of specialist fields, where statistical analysis of results from trials are needed and thus standardised medical records are already devised and used during a trial. Therefore investigations can be undertaken on what has actually been done for and to patients and, above all, what the outcomes were (Konner, 1993).

However, to ensure that the importance of the results from these investigations can be quickly and easily identified and understood by the clinician, they need to be presented in an appropriate format. This will help to capitalise on one activity in which humans are superior to current computer systems, i.e. the ability to deal with data flexibly. 'We automatically notice unusual patterns and derive general rules of thumb without conscious effort' (Parsaye and Chignell, 1993). However, this can only occur if the information is present in such a manner as it can be 'chunked' and integrated with knowledge investigators already know. The new information and knowledge that has been gained through analysing the results of the trial can then be used to initiate further trials and hence a clearer understanding of the domain and of the decision process will evolve.

These types of investigations are important within many specialist medical fields since currently there is 'no formal language [that] has been developed yet but a fairly stable, well-agreed language does exist, (for example, in diagnosis and medicine)' (Kidd, 1987). This lack of formalisation within a field therefore causes major difficulties if an attempt is made to automate the decision-making process used. The problems inherent in trying to define human decision-making process are described in Chapter 5.

However, if the actions and language within a domain are already formalised, computer systems can be built which can mimic these defined procedures. In these situations where humans are merely applying established rules and set methods 'machines are better than people: they are faster, more efficient and more reliable. However we must remember that these domains concern the less human part of our intelligence, and do not include intuition, motivation, judgement, ethics and wisdom. We need these aspects of our intelligence at all times because any formalised model is an incomplete description of reality' (Gremy, 1989). Pauker et al (1976) agree, that 'formalisms alone e.g. cybernetics, mathematical logics and information theory cannot produce intelligent behaviour in complex "real-world" situations. It has become abundantly clear that no simple formal approach can accommodate the knowledge of first principles and experience, common-sense and guesswork required for "intelligent" activities.' Therefore computers are best suited to those situations which can be formalised. In this context accuracy, speed and reliability are paramount.

Humans excel, however, in reacting to "real- world" situations which require numerous different kinds of knowledge and experiences to be applied in an ad-hoc fashion. Consequently, where no formal language exists as yet, investigations should be undertaken to attempt to identify those areas within the domain that can be standardised - where rules and procedures can be established for the specific domain and the evolution of a formal language can be initiated. To achieve this, facts regarding situations within the specialist domain must be gathered, the outcomes of actions understood, and ultimately proposed theories and hypotheses proven through the investigations of past events. Therefore, methods by which facts within the specialist area can be collected, stored and analysed, must be developed. These are discussed in Chapters 3 and 4.

Currently, however, to be able to develop an appropriate computer system, users must possess a substantial amount of computing and computer knowledge. Yet within many specialist domains people have expertise in areas other than computing but appreciate the benefits that computerisation might bring to their own domains. Unfortunately the systems that they require are generally relatively simple although they do require tailoring to the specialist environment in which they are to operate. Without the necessary computing knowledge, the prospective users can not construct their own systems using the available packages and as the systems often do not hold enough unique or challenging qualities, researchers tend to be disinterested in developing such systems. Furthermore, hiring a computer programmer is often fraught with problems and is very expensive. Therefore, the benefits of such systems tend to be lost. One method of rectifying this situation is to develop appropriate tools which will enable the domain experts to construct their own systems.

The special considerations and requirements for naive computer users and for implementing systems within medical domains are discussed in more detail in Chapters 3 and 4. One technique that can be used to help overcome the fears of inexperienced computer users and to aid their acceptance of the computer system, is to include the users in the design and development stages of the end-system. By this approach the users are "trained" as the end-system itself develops. Another method is to introduce users to relatively simple systems, systems that assist with routine mundane tasks and do not threaten the users' position or authority. In general, once users are confident with such systems, they will request extensions and additional facilities to achieve other tasks. Consequently, it will be the users who will be driving and dictating the development of the computer system - it will be their system.

Only with the co-operation of the prospective users will a successful system be built and only with their commitment to and their understanding of the computer system, can it be then used effectively. Without the enthusiasm of the users there is little prospect of the benefits of computerisation being realised, if in fact a fully operational system could ever be constructed and implemented at all: 'The success of a decision support system will be measured by its ability to make its users think more and better i.e. *(sic)* to use their intelligence to its highest potential', (Gremy, 1989).

This research is proposing a methodology which will enable naive computer users, who are experts within a specialist field (i.e. a medical consultant), to design and build their own Intelligent Database Decision Aid (IDDA) by using a suite of computer-based tools. Once the tailor-made IDDA has been developed by the domain expert, it will provide the facilities to collect and retrieve information. In addition, it will enable statistical analyses to be carried out on the stored data, thereby providing information to assist in the making of more informed and, hence better, decisions. Moreover, for those physicians who conduct research, such a database system would ease considerably the tasks of data storage, review, and analysis, which are essential for such an activity. For those who do not wish to undertake original research, there will be the ability to replicate published studies to determine whether the findings can be generalised to their own patient populations. The following section briefly describes the background to this research.

1.5 Background

This research evolved from a previous project with the orthopaedic division of the Leicester Royal Infirmary. That study involved investigating computer-based decision support systems for knee ligament injuries. A manual assessment system already existed at the hospital. This method, however, was becoming unworkable due to the increasing amounts of collected data. Even the simplest reviews or analyses were taking so much time and effort on the part of the doctors, that the whole assessment process was under threat. Consequently, the decision was made to computerise, and a project was initiated. Investigations into the manual procedures were undertaken and a model of the assessment process was established. Due to the financial constraints of the NHS, the system had to be PC-based and the software selected was dBASE III+.

By the end of the one year project, appropriate databases had been established and interfaces for the data entry and the viewing of records had been constructed. In addition, a number of standard bar charts, which had been defined by the doctors, had been implemented (Jackson, 1988). The procedures and assessments modelled by the computer system matched those of the old manual system. This was to aid the transition from the manual process to the computerised method, therefore assisting user acceptability of the end-system.

This system has now been in operation at the Leicester Royal Infirmary (LRI) for the past six years. It was from this basis that this research evolved. However, that initial system too has developed with the progression of this research. For example, users can now control and define the bar charts to be displayed. Furthermore, various statistical analyses can be carried out on data selected by users and the results presented in an appropriate manner. These extensions are a direct result of the evolution of the methodology for the current research. In addition, feedback from the medical consultants regarding the LRI system, influenced the research developments. The establishment of this two way flow of information and theories was therefore an important and influential factor in the progression of both the research methodology and the LRI system.

Moreover, it has become evident that such a system was not merely a one-off request from an individual medical specialist. Other projects similar to the knee ligament system have been requested from hospitals in the Leicester area. These deal with other orthopaedic specialisms, for example, knee replacements and hip replacements. Hence, the indications are that medical specialists are now wishing to explore the possibilities of computeristation and are requesting applications designed specifically for their own particular domains.

A number of these liaisons was used as pilot studies for this research. Thus, during the development of the methodology, various scenarios were examined to determine whether the proposed technique was flexible enough, and indeed feasible to use, in the production of appropriate end-systems for different specialist fields. Hence, a build-test-review cycle was used during the development and progression of this research study.

At the outset, however, initial investigations were undertaken to discover whether a standard system was already in use within specialist fields or, if no standard system existed, what different types of system were in operation or indeed if any were being used at all. A number of international orthopaedic institutions was selected from the medical literature. They were chosen because their papers, or the presentation of their results, indicated that they may have used a computer to collect and/or analyse their data. Of the eleven institutions who replied, over half had either not used a computer, or else had merely carried out statistical analyses on a set of data entered solely for the investigation under review at that time. In three of the other four cases, statistical advisers were handed the information to carry out all the analyses. Only in one institution, Cincinnati Sports Medicine, had a computer system been designed and constructed to record all the patient details and their past histories. Here a Compaq Deskpro 386 was used and the software was Rbase system V. A professional computer programmer had been employed to develop the system, taking two years to arrive at the point where the doctors could review and analyse the data. The doctors believe that they are already seeing the benefits of this system - 'we now feel that with the use of our knee rating system on the computerised database that we now save hundreds of hours and dollars in being able to perform all of the statistical analyses in-house.'

If this freedom is extended even further, enabling doctors themselves to construct their own computer system as and when required, the benefits of computerisation could become even more apparent. For example, there would be no reliance on an outside contractor, no extra expense everytime a new system is required; and, no time delay caused when a system is being constructed to someone-else's timescales. However, the currently available computer packages and tools require their users to possess a substantial background in computers and computing. It is rare that medical specialists have such knowledge. Moreover, they have little free time in which to acquire the appropriate skills. Therefore it has been necessary up to now to enlist outside expertise and assistance, even though there are often drawbacks in this approach, as has already been explained.

This research, however, is proposing a standard methodology which encompasses all of the processes and procedures involved in designing and building such systems and a set of computer tools which implement this technique. Furthermore, this approach has been devised to allow experts to interact directly with a computer to develop their own computerised systems. This enables an expert to dictate and tailor a system to the domain specific requirements. As the investigations undertaken have indicated, there is a lack of such techniques or appropriate tools for inexperienced computer users to operate. In addition, such systems which assist the expert, even in a specialist field, are notable by their absence. Even so, doctors are beginning to realise the benefits that computers can bring, and are now becoming more willing to co-operate. Consequently, there is a need for such resources. This research is therefore proposing such a methodology which will enable effective systems for specialist fields to be constructed quickly, cheaply, and easily by end-users, who, themselves, may well be naive computer operators.

The first process is to understand the requirements involved when undertaking an investigation within a specialist medical field. Previously expert systems attempted to acquire from a medical expert, the underlying knowledge of decision-making processes and then model these. However, the structure of the human brain, the methods used by humans to store information, the different type of knowledge utilised by humans and the nature of 'expertise', were all influential in restricting the access to the crucial pieces of information. The next chapter reviews those problems inherent in human memory, expertise and knowledge acquisition, and summarises a number of techniques currently used when attempting to obtain the specialist domain knowledge. It concludes by describing and explaining the approach adopted in this study.

Chapter 2

Memory, Expertise and Knowledge Acquisition

2.1 Introduction

The focus of traditional AI has been in domains which can be isolated from the influences of the 'real-world'. However, humans do not operate in such an enclosed setting. They are social beings, who interact with others, with the environement and with themselves. Even cognitive science has been criticised for ignoring these aspects of behaviour (Norman, 1990). Instead there has been the perception that humans consist of pure intellect, 'communicating with one another in logical dialogue, perceiving, remembering, thinking where appropriate, reasoning its way through the well-formed problems that are encountered in the day' (Norman, 1990). This however, is not how humans actually behave.

Moreover, humans are individuals and, though they are likely to use similar storage and retrieval mechanisms, the different types and items of knowledge, data and relationships which exist within one person's memory will be unique to that person. 'All organisms achieve some presentation of their environments adequate for their survival as a species, although they do it in very different ways' (Miller,1990).

Therefore to model, in a computer system, expert behaviour within a domain, the underlying knowledge and the nature of expertise must first be established. This involves knowledge elicitation and acquisition. Knowledge elicitation is the process of creating a representation of an expert's (or experts') competence in a field of activity. Acquisition is the process of collecting the detailed information which fit into the model (Addis, 1987; Birmingham and Klinker, 1993).

An expert system's performance is dependent upon the quality and completeness of its embedded knowledge. Hence extracting and formalising the domain knowledge has often been described as the most critical stage in the expert system development process (for example, Diaper (1989)). However this transfer of expertise and knowledge from a human source to a computer is complex and poorly understood. 'We cannot be satisfied until we have a precise understanding of the processes underlying expertise, its operation, acquisition and transfer. This is not a simple requirement since it entails understanding the nature of knowledge, its dynamics and application. The foundation of computer technology in the physical sciences is no longer adequate and need extension into the humanities. The philosophy, psychology and sociology of knowledge processes are highly significant to future computing and we have to make theories operational and obtain precise answers to questions that have long been regarded as certainly controversial and possibly intractable' (Gaines,1988).

The following chapter discusses a number of the key reasons why the tasks of knowledge acquisition and transfer have proved to be so difficult to accomplish. First, however, there is a description of the structure of human memory, followed by a brief review of a sample of different types of knowledge utilised by humans. The differences between an expert and a novice are then explored. From these summaries the kind of problems knowledge engineers face in the knowledge acquisition process will then become apparent. These are discussed further and a number of the most commonly used tools and techniques are reviewed. Finally, the methods proposed for use in this research are described and explained.

2.2 Human memory

The human memory enables a person to learn - 'If we remembered nothing from our experiences we could learn nothing. Life would consist of momentary experiences that had little relation to one another. We could not even carry on a simple conversation. To communicate, you must remember the thoughts you want to express as well as what has just been said to you' (Aitkinson et al, 1983).

Psychologists have made two distinctions about human memory. The first concerns the three stages of memory - encoding, storage and retrieval. The second deals with the two kinds of memory - short-term and long-term. The basic structure, function, organisation and the differences between these two types of memory and how they relate and interact must be understood if a computer system is to be built that can mimic a human with any degree of success.

Atkinson and Shiffron (1971,1977) proposed one theory, the Dual-memory theory, to explain the relationship between short-term and long-term memory (figure 2.1). This theory assumes that information enters the short-term memory where it can either be maintained by rehearsal or lost by displacement. In order for information to be encoded into long-term memory it must be transferred there from short-term memory. Therefore all items encoded in a person's long-term memory must first have been stored by their short-term memory.

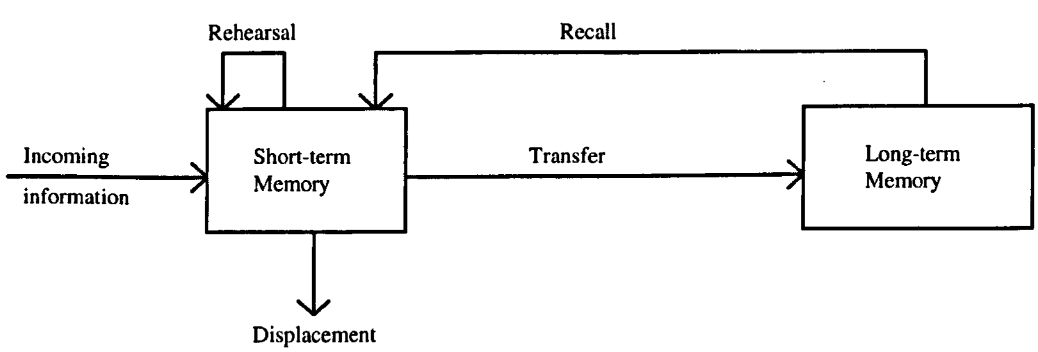


Figure 2.1: Dual Memory Theory (in Atkinson et al, 1983)

Short-term memory has a very limited capacity approximately seven items, give or take two (Miller, 1956). Long-term memory, however, seems to be unlimited. There are other differences too (figure 2.2). These include - the method by which items are encoded (the short-term memory favours an acoustic code whereas with long-term memory it is based on meaning) and the difference in the ability to retrieve items (short-term memory is thought to be more or less error-free while long-term memory appears to be error-prone).

Atkinson et al (1983) propose that many cases of forgetting items from long-term memory seem to result from a loss of access to the information rather than from a loss of the information itself. Therefore a poor memory may reflect a retrieval failure rather than a storage failure. This is unlike short-term memory where forgetting seems to result from exceeding storage capacity rather than through problems of retrieval which is thought to be virtually error-free.

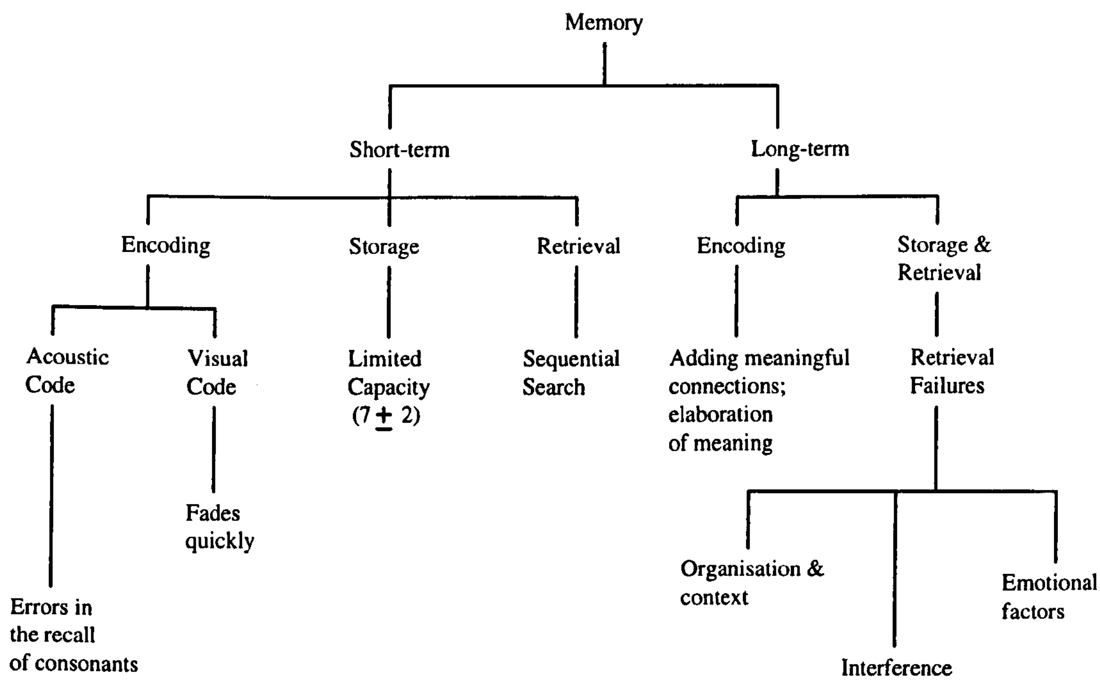


Figure 2.2: Memory Organisation (in Atkinson et al, 1983)

Organisation improves retrieval, presumably by making the memory search more efficient. The ability to retrieve items effects problem-solving. Consequently, memory organisation will definitely have an effect on decision-making.

2.2.1 Structure of long-term memory

Facts can be gathered by various means, generally by being seen, heard or sensed in some sort of way. People however do not "store" facts in their long-term memory. Instead items are retained by their meaning. Knowledge is based on this semantic representation of facts and the interrelationships that exists with other information (Addis,1985). Humans are thus good at recalling the meaning of information, i.e. a passage of text, whilst being extremely poor at remembering exactly how the material was originally presented (Thomson,1984; Sandberg and Weilinga, 1991).

Psychologists have attempted to establish the basic structure of the human memory and how humans retain and process knowledge and information. The three main theories that have evolved are - 'categorisation' (Bergson, 1911; Rosch, 1978; Minsky, 1985), 'conceptualisation' (Piaget, 1950; Sowa, 1984; Berg-Cross and Price, 1989) and the 'Personal Construct Theory' (Kelly, 1955; Shaw, 1980; Garg-Janardan and Salvendy, 1987).

All are based on the same belief that long-term memory is highly organised and is built up of hierarchical structures of information linked together by their common associations. The following two simple descriptions of categorisation show the necessity for the existence of some type of classification structure - 'categorisation is a basic process in the construction of any representation: at the very least, substances must be categorised as edible or inedible and organisms must be categorised as friend or foe' (Miller,1990). 'We are able to identify objects and events by placing them quickly into preconceived categories; this reduces the strain on our nervous system by rendering recognition or identification automatic and by reducing the amount of learning we have to tackle. If we didn't categorise or classify automatically we would be faced with the exhausting and complicated task of relating every particular item in our experience to every other item in the context of their occurrence' (Thomson,1966).

A purely hierarchical structure would not however account for the diverse behaviour that people exhibit under the same conditions. Kelly (1955) explained this phenomenon by claiming that people build their own version of reality and that the hierarchical structures of personal constructs have associated values attached to them by the person. He believed that these values effect how a person behaves by acting 'like a pair of spectacles colouring and focusing a person's internal and external worlds' (Shaw, 1980).

With the theory of concepts, two uses were defined - the extensional use and the intensional use. The extensional use, which is more or less the same for everybody, indicates the actual object which the concept denotes in a direct manner, i.e. this one (pointing to it) is a chair and that one is a stool. The intensional use however can vary from person to person and is derived from the experiences of that person, i.e. the concept 'dog' can cause terror in X, joy in dog-lover Y and make Z, a veterinary surgeon, think in a scientific manner. All have an identical extensional use of the concept: applying 'dog' to one type of small quadruped. However, intensionally the concept has a different meaning for each. It is claimed that almost every concept has some intensional use determined entirely by the experiences and make-up of the individual who uses it (Thomson, 1966).

Though each approach may differ in the manner it proposes for the structure of long-term memory, they all agree that the knowledge stored cannot be seperated from its original context, i.e. it is essentially situated. This context, in turn provides a basis for thinking about, and reasoning with, the knowledge (Davis and Bostrom, 1993). Moreover, not only is it embedded in a particular frame of reference, knowledge is also subjective and open to interpretation. This, therefore, makes knowledge transfer extremely difficult. 'Each act of speaking is a complete act of perceiving in itself. By speaking we create new meanings which are perceivable by ourselves and others ...... as we can only know the things we perceive, it follows that all knowledge is relative to the observer. Our perception is biased and perceiving implies conceptualising, interpreting. What we perceive is an interpretation in itself' (Sandberg and Weilinga, 1991).Therefore, experiences and knowledge are constantly being organised and reorganised on the basis of new and/or similar experiences, knowledge and cultural norms.

Miller (1956) described the method by which incoming information is grouped and linked together as "chunks", e.g. clusters of symbols associated with a set or pattern of stimuli. Zhang (1988) explains with an example :- 'when children start to learn about a man, they may begin to organise their experience into chunks (figure 2.3). As they become older, more and more chunks are added to their memories and hierarchically organised by their relationships. Moreover, if one becomes a medical student many more chunks are added to his/her memory and clustered together around successively more abstract concepts.'

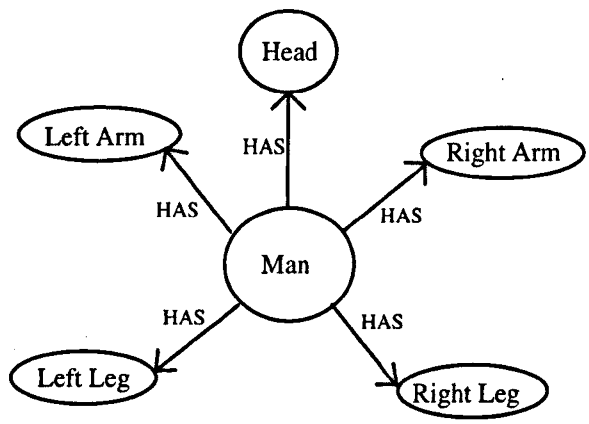


Figure 2.3: Chunks of a man

Schank (1990) calls these groups of information "memory organised packets" (MOPs). They are basically the same - the method by which experiences and information, gathered from different episodes, are organised into sensible units around essential similarities. Their purpose is to enable associations to be drawn between previously encountered structurally similar experiences or information and newly acquired items of knowledge. This enables the prediction of future events and allows implicit links to be drawn between items of information, i.e. having previously acquired the characteristics of the concept 'dog', a person who has never seen a 'Beagle' but is told it is a type of 'dog' will automatically link 'Beagle' with their own known 'dog' characteristics.

2.2.2 Different types of knowledge

Not only is the precise structure and organisation of long-term memory unclear but so too are the types of knowledge that humans access and process. Klein et al (1989) defines the three classes of knowledge necessary for expertise as being -

a) explicit and objective knowledge, i.e. factual knowledge, if/then rules and analytical procedures,
b) tacit knowledge (which is hard to articulate), i.e. contextual knowledge, analogical inferences and judgments of typicality,
c) perceptual learning and the development of perceptual-motor feel - skills are mastered, finer discriminations are made and tools come to be manipulated automatically, e.g. driving a car requires perceptual learning.

Psychological research (Anderson, 1982; Gordon, 1989; Rumelhart and Norman, 1990) seem to indicate that there are two different kinds of knowledge that people use when they perform any kind of cognitive task. These are declarative knowledge and procedural knowledge. Gordon (1989) describes both, 'declarative knowledge consists of what we know about objects, events, static relationships between concepts and so forth. --- Procedural knowledge is more difficult to describe. It is essentially the knowledge about how to perform various cognitive activities, the dynamic process of operating upon knowledge.' Declarative knowledge tends to be accessible, thereby enabling it to be examined and combined with other declarative statements to form an inference. Procedural knowledge however tends to be inaccessible. As Rumelhart and Norman (1990) explain, 'although we can pronounce a word like 'serendipitous', we cannot say what movements our tongue takes during the pronunciation without actually doing the task and noting the tongue movements'. Hence though there does seem to be conscious access to declarative knowledge, this access to procedural knowledge does not seem to exist.

Anderson (1982) in his model of learning demonstrated how declarative knowledge can be transformed into procedural knowledge. This occurs when a piece of knowledge is used over and over again in a procedure which results in the access to it being lost and thus the ability to report it verbally is also lost. Berry (1987) defines this as the first type of implicit knowledge - that which was once represented declaratively or explicity.

The second form of implicit knowledge which Berry (1987) describes, arises from the autonomous stage of the learning cycle, i.e. the knowledge that was gained as the result of an implicit learning process. It has never been represented explicitly or declaratively - 'as far as human experts are concerned they not only have difficulty in describing what they do because their knowledge is no longer in declarative form but because some aspects of their knowledge have never been represented explicitly. They have learned through experience rather than being picked up from one or more textbooks' (Berry,1987).

Gruber (1988, 1991) identifies yet another kind of knowledge - strategic knowledge. It is used to decide what course of action to take when there are conflicting criteria to satisfy and when the precise effects of actions cannot be known in advance. An example is driving through a city, an action that requires strategic knowledge. The driver has to choose actions to balance criteria such as minimising travel time, maximising safety and driving pleasure. Strategic knowledge is intrinsic to tasks in which managing the process of problem solving, choosing among possible actions, is part of the expertise. 'Strategic knowledge is about how to play rather than the rules of the game' (Gruber, 1988).

Physicians often make strategic decisions in their jobs. For example when choosing diagnostic tests: they consider the possible effects of their actions on multiple criteria such as patient safety and comfort, the economic cost, protection against missing a dangerous condition and of course increasing the likelihood of making a correct diagnosis.

These are only a few of the various kinds of knowledge that a person possesses. The actions and the decisions reached by an individual are dependent upon the actual types of knowledge utilised for the particular problem in hand. However the type of knowledge used is not the only factor.

Beliefs and biases further influence how a person acts or reacts in a situation. 'Beliefs are picked up like fashions, through the unconscious or half conscious or deliberate imitation of examples set by some group of people' (Thomson, 1966). They are formulated because of certain experiences in the past: 'a man's most fixed beliefs carry the stamp of his personal history' (Thomson,1966).

Therefore, people can be unaware that they follow or maintain certain beliefs. This situation can persist until some crisis or argument forces them to reflect, acknowledge and define their beliefs. This is a major problem in knowledge acquisition since forcing experts to review and acknowledge their beliefs can undermine the confidence of the experts. Even so, these same beliefs effect how a person will act in specific situations by influencing how a person perceives and views the available evidence. As Simon (1990) explains, 'in the case of problems of types that he encountered previously, the understanding process may be determined by that previous experience and may be different for different subjects'. These beliefs and biases must therefore be defined if a true presentation of a person's knowledge base is to be modelled.

The differences in the behaviour and actions of people can also be attributed to the inaccuracies in the actual knowledge retained by a person. Inaccuracies can occur during the communication of information since the general method involves the use of vocabulary, whether it is read or spoken. The abstract nature of vocabulary itself, can result in a message being altered extensively, since in order to retain a message people impose their own meaning on it.

The weightings and biases, created when a person interprets and reorganises knowledge to permit retention, have serious implications for those attempting to acquire knowledge since the knowledge elicited may lack consistency and accuracy due to it having been falsely maintained by the person. In addition, these inaccuracies in the underlying knowledge will affect problem solving and decision making since these processes are reliant upon the knowledge that has been retained. The biases will determine how the person views the available evidence and therefore influences what decision is reached and how this is achieved (Evans, 1987b).

People can therefore draw different conclusions even though they have been subjected to the same evidence. 'For example, it is not unusual for scientists to construct rival accounts of phenomena which they maintain for many years, despite being exposed to the same evidence - the research literature - as their opponents. Nor is it uncommon for the experiments [to be subconsciously] designed in particular laboratories to maximise the chance of confirming the theory which motivated them' (Evans, 1987a).

2.3 Expertise

The major difference between an expert and a novice, it seems, is in the perception of the information and the ability to efficiently organise and categorise information and remember rules that link chunks together (Mayer, 1981; Olson and Rueter, 1987). This results in the expert being able to recall, at an instant, much more information that is relevant to a task than a novice would be capable of. Whereas novices seem to solve problems by using declarative knowledge and considerable conscious control over the problem solving process, experts on the other hand seem to use automated sequences which are not controlled consciously (Mayer, 1981; Davis, 1989a). Therefore 'instead of perceiving and remembering individual pieces of information, experts process meaningful chunks, making their perception more efficient and thus their recall performance higher than a non-experts' (Jefferies et al,1980).

Klein et al (1989) discovered from their research into the development of a knowledge elicitation strategy, that experts rarely consider more than one option during problem solving. An expert's ability to handle decision points appeared to depend upon skills at recognising situations as typical and/or familiar. It is this recognition that suggests feasible goals, alerts decision-makers to important cues, suggests promising courses of action and generates expectancies.

Pauker et al (1976) had observed similar behaviour during their research. Furthermore they identified how an expert reacted when an initial hypothesis proved to be incorrect - 'he [the expert] often employs the rather efficient strategy of associating one hypothesis with others with which it may be readily confused (e.g. "multiple pulmonary emboli are often confused with cardiomyopathy"). By explicitly remembering such situations, the physician can move directly from an hypothesis which has become suspect to one which offers another plausible explanation for the presenting findings.' They discovered a novice however did not react in this manner - 'the medical student or young physician does not have an extensive knowledge of such relations and so is unlikely to move from one hypothesis to another in such a skillful fashion.' Therefore they observed that students, to counter this problem, would approach the diagnostic process in a highly structured, methodical manner. This involved collecting more information, and in greater detail, before the selection of a hypothesis was made, thus reducing the likelihood of serious errors.

These findings could be seen to support Bartlett's (1958) belief that the other major difference between the performance of an expert and a non-expert was timing. Experts seemed to have lots of time. They did their tasks easily, smoothly, without apparent effort and with plenty of excess time - 'experts are not only proficient but also smooth in the actions they take' (Johnson,1983). The non-expert however is always in a hurry, barely able to cope, and rushing from this to that. With the non-expert, the difficult task looks difficult.

This observable difference could also result from the ability of experts to use domain knowledge flexibly depending upon the immediate goal, e.g. experts apply their wealth of knowledge in a variety of ways to various activities, i.e. problem-solving, teaching, explaining and finding inconsistencies in reasoning patterns (Wielinga and Breuker, 1987). Hence experts structure the different tasks in a manner that makes the most effective use of their knowledge and they use different strategies for different types of problems. With this approach, therefore there is a 'graceful degradation' when a human expert is confronted with a 'new' problem.

Consequently, expertise does not seem to rely solely upon possessing a large knowledge base but also requires the ability to efficiently organise and categorise associated information and experiences and to recognise and link previous similar events with the current situation under review.

Kolodner (1983) agrees, believing that knowledge is built up incrementally on the basis of experience, thus enabling facts which were once unrelated to become integrated. Gradually the reasoning processes are refined, and usefulness and rigidity of rules is learned. Therefore, the evolution from novice to expert requires introspection and examination of the knowledge used during problem solving. This, in turn, implies that as experts are having new experiences, they are evaluating and understanding them in terms of previous cases. During this process, they must also be integrating the new experience into their memory so that it too will be accessible for use in the understanding of any subsequent problems. This cycle continues, thereby enabling experts to regularly update their knowledge base as their experiences evolve.

Anderson (1982) divided this learning process into three stages -

a) Cognitive stage: a person learns from instructions or observations the appropriate actions for specific situations,

b) Associative stage: the stage 1 relationships are practised until they become smooth and accurate,

c) Autonomous stage: the relationships are compiled through practice to the point where they can be done 'without thinking'.

Norman (1990) gives an example of this cycle - 'Consider what happens when you first learn to drive an automobile. The instructions you receive emphasise the actions and mechanisms: hold the steering wheel this way, synchronise foot (for clutch) and hand (for gearshift) that way. As you progress, the point of view changes. Now you are turning the wheel, not moving your hands clockwise. At the expert level, you may no longer be aware of all the subsidiary operations that you perform: you look at the driveway, form the intention to enter, and the car obediently follows suit. Driving becomes as natural as walking, the car becoming as much a part of the body's appendages as the limbs.'

This process of compiling relevant relationships and actions holds further problems for anyone attempting to emulate or mimic expert behaviour since the experts themselves have become unaware of the importance of such knowledge (Shiffron, 1977). Collins et al (1985) and Berry (1987) argue that certain actions and knowledge, essential for success, may never have been consciously known to the expert. Furthermore, Mouradian (1990) claims that certain types of knowledge used may never be described verbally - 'details are fleeting and observed in passing, and many of the observations are of nonverbal behaviour. As a result, the clinician will have difficulty in articulating the details of their *(sic)* observations and reasoning, remembering only the overall impressions. Decision making in medicine is quite intuitive.'

A major part of expert problem solving is therefore carried out subconsciously. This causes problems during knowledge transfer since experts cannot articulate associations and information that are linked to their activities, if they are unaware of their significance.

2.4 Knowledge elicitation, acquisition and related problems

Knowledge elicitation and acquisition are the first major hurdles in the construction of any expert system. The time required for this stage influences greatly the development costs and performance of the end-system. 'Expert knowledge is not easily captured. The process is time-consuming, painstaking and complicated' (LaFrance, 1987). Reliability, validity and the usefulness of expert systems are reliant upon the knowledge acquired since this is the foundation upon which everything else is built.

Shortliffe et al (1984) described knowledge acquisition as 'the transfer and transformation of problem-solving expertise from some source to a program'. Kidd (1987) explains what is involved -

'1) Employing a technique to elicit data (usually verbally) from the expert

2) Interpreting these verbal data (more or less skillfully) in order to infer what might be the expert's underlying knowledge and reasoning processes

3) Using this interpretation to guide the construction of some model or language that describes (more or less) accurately the expert's knowledge and performance. Interpretation of further data is guided in turn by this evolving model'.

The source is generally a human expert but it could be original material, i.e. journal articles, text books or experimental data. The type of knowledge that is required is a collection of definitions, relations, specialised facts, algorithms, strategies and heuristics about the narrow domain area. Therefore, both declarative and procedural knowledge must be acquired.

Keyes (1990) believes that though approximately 80% of the knowledge in a given field can be gleaned from text books and procedure manuals, it is the other 20% 'the heuristics, the gut feeling, the "what makes this person special" that is so hard to encode'.

Mouradian (1990) agrees that journals, manuals and results from experimental data do not reveal all the knowledge. Furthermore, the translation and standardisation of the information to remove biases and deviations in practices and the interpretation of results are not only time-consuming but are also extremely difficult to carry out: 'building a knowledge base from the scientific literature is likely to be frustrating. Estimates suggest that less than 1% of the articles in some specialities fulfil scientific criteria' (Mouradian, 1990). He continues by pointing out that the literature, or the lack of it, might itself be misleading - 'Any scientific discipline relies heavily on early literature to clear avenues for research, but in medicine more trends disappear than remain. However, the literature rarely reflects the fading trends, thereby leaving a peculiar literary-historical bias. Only favourable articles may exist about a given procedure although its popularity may have long since waned. Therefore, should you assume a decline in popularity from a lack of recent literature? You may as easily make the reverse error by concluding that a procedure is no longer popular because no new literature is available - which, of course, is not necessarily true' (Mouradian, 1990).

Chalmers et al (1990) agrees, stating that abnormally good results are much more likely to be published than poor ones. Consequently it has mainly been human experts who have been relied upon to divulge the necessary knowledge and their underlying expertise and experiences in the relevant field - 'knowledge elicitation from humans is the major thrust here since the intuition, experience and heuristics used by humans in problem-solving are rarely stated explicitly in the literature' (Garg-Janardan and Salvendy, 1987).

Unfortunately it is not easy. Knowledge cannot be 'mined out' of an expert's head like nuggets one-by-one as Feigenbaum and McCorduck (1983) imply. Nor, as LaFrance (1987) correctly points out, is it just a case of simply having experts 'hooked-up and drained of everything they know.'

Knowledge about cognition is mainly acquired from experts by having them discuss and reflect on their own problem solving processes. Though these reflections are important they often result in general statements of principle rather than revealing insights into the precise strategies used or how these strategies are applied in particular cases. The problems arise because certain kinds of knowledge are not easily communicated due to their psychological form (e.g. mental images) or the inaccessibility of the knowledge (a lot of reasoning takes place below the level of conscious awareness) (Collins et al, 1985; Davies and Hakiel, 1988; Diaper, 1989). The outcome is that there may be little correlation between verbal comments made by an expert and the real items of knowledge that are actually used (Kuipers and Kassirer, 1984; Gremy, 1989; Byrd et al, 1992; Guggenheim and Whitfield, 1991). As Thomson (1966) states, 'a person may respond habitually and consistently towards a particular type of stimulus situation without being able to discriminate and describe what he does or what his motive is for doing it.'

Therefore situations occur when the information given as an explanation for an action is in fact incorrect. For example, Collste (1992) discovered that an original description did not respond to the actual way a doctor had acted when he was confronted with the problem in a real setting: 'When this was pointed out to him, the doctor answered: "Oh, I know that, but you see I don't know how I do diagnosis, and yet I need things to teach students. I create what I think of as plausible means for doing tasks and I hope students will be able to convert them into effective ones".'

Nisbett and Wilson (1977) reported on several cases when this phenomenon occurred - where people seemed to be unaware of the stimulus factors that determined their responses. They concluded consequently that only the 'product' of the mental process was accessible and could be ascertained via verbal reports whereas the 'process' by which the choices were made during decision-making could not.

The problems are further compounded when multiple experts are required to participate in the construction of one expert system since there is often conflict between the knowledge acquired from different experts as well as deviations in the emphasis and importance of particular facts and procedures carried out by individuals. As Addis (1987) states - 'an expert's perception of a task domain may be determined by the part which he plays within the task domain, the functions he carries out and the interactions he experiences. In this way the same task domain could be described very differently by two individuals who, due to different objectives and requirements, see different facets of the domain as being important.'

Consequently there should be little surprise over reports of experts in the same field, faced with the same problem, acting in different ways and reaching different decisions (Evans, 1987a; Fieschi, 1990). One question for knowledge engineers is 'Who is right ?'. If on the other hand an expert system is built purely around one expert, the acceptability of that system will be greatly reduced, not only by other experts in that field but also by those less experienced. Consequently, the result has been that decision support systems are only used by the people, who actually developed them (Hripcsak, 1991).

2.4.1 The knowledge engineer

Knowledge engineers are therefore faced with a number of difficulties during knowledge acquisition.

Firstly, the problems, already discussed, concerning the difficulties that experts experience when attempting to describe the methods they use to accomplish tasks. Furthermore, this can lead to an expert feeling uneasy about being questioned 'because [when] experts cannot come up with answers to what seem like reasonable questions they can end up feeling threatened by the knowledge elicitation process. This will result in their seeing themselves as being inconsistent or illogical thinkers' (Berry, 1987).

Secondly, as the majority of knowledge engineers use interviews at various stages in the knowledge acquisition process, there can be human biases in the judgments and the transferred knowledge introduced both on the part of the expert and the knowledge engineer - 'It is well known that verbal data can be interpreted in a variety of ways, depending upon the viewpoint of the speaker and listener, the assumed background knowledge and the possible social effects' (Wielinga and Breuker, 1985).

In addition, experts may not articulate all of the required knowledge since some of the information may be regarded by the expert as being commonsense (Fox, 1984). However, knowledge engineers may well be unaware of its existence. This occurs not only because knowledge engineers lack the necessary expertise in the field but also because experts and knowledge engineers have different experiences and goals as well as preferred ways of thinking and undertaking problem-solving (LaFrance, 1987; Silverman, 1990; Mykytyn et al, 1993). This can lead them therefore to follow different thought processes and strategies to arrive at the same final conclusion. Consequently, the knowledge engineer may be building up incorrect inferences based on incomplete or inaccurate information and beliefs.

Finally, though possibly most importantly, knowledge engineers must be well-versed in a problem domain before starting knowledge acquisition with an expert. This will help to counter the biases created by misinterpretations and to help spot inaccuracies and inconsistencies in the acquired knowledge.

Furthermore, it will enable knowledge engineers to understand the terminology and structure of the subject area to produce meaningful and relevant questions (Birmingham and Klinker, 1993; Silverman, 1990; Byrd et al, 1992; Keyes, 1990). Therefore, all conversations tend to be carried out in the terminology and language of the expert. This is not only to aid in the initial stages of knowledge acquisition but also permits the expert to assist in the crucial stages of refining and evaluating the knowledge of the end-system. 'If the domain expert [DE] does not understand the language used for KB [knowledge base] construction and the processing mechanisms, and the KE [knowledge engineer] is not knowledgeable as the DE, the KE may have a difficult time finding some method for validating proper knowledgebase construction' (Lynn and Bockanic, 1994).

Collins et al (1985) place so much importance on the knowledge engineer being competent in the domain that they suggest that an apprenticeship is undertaken within the problem area - 'the knowledge engineer must do more than tap the knowledge of the expert at one remove, but must undertake at least a short apprenticeship - a period of participant observation - as part of the elicitation process.'

Berry (1987), however, believes that this is not a suitable or practical solution to the knowledge elicitation problem and that it is merely transferring the problem. The general consensus is therefore that though knowledge engineers need not be classified as experts, they must be able to discuss competently with domain experts the whole problem area and the various aspects of their expertise.

Klein et al (1989) stress the necessity for understanding and appreciating the various elements of proficient performance. Since any knowledge elicitation technique emphasises some aspects of expertise and de-emphasises others, they suggest that knowledge of expert performance is first established before selecting an acquisition technique. Knowledge engineers must therefore understand both what is being captured and what is being missed to enable a full scenario of the performance in a task to be developed (Birmingham and Klinker, 1993; Brooking, 1986).

This results in knowledge engineers being faced with an incredibly steep learning curve at the start of any project. Furthermore whilst this learning process proceeds, the actual development of an expert system cannot progress, since the rest of the system is built on or around the knowledge gathered at this initial stage. To start too soon will almost always require extensive rebuilding and/or redesigning. Start too late and the cost effectiveness of the system will shrink or disappear altogether.

This time requirement in itself is yet another problem that the knowledge engineer must resolve. Generally experts cannot afford, or are unable to give up, enormous amounts of their time to build an expert system. Knowledge engineers are therefore faced with having to solve the difficulty of the 'limited availability of experts in disciplines where the expert is unique or indispensable and cannot be spared from the day-to-day task' (Grover, 1983). Consequently, the time that an expert can spend with a knowledge engineer must be used to the full and as much knowledge must be obtained as efficiently and effectively as possible.

The selection of appropriate tools and techniques to acquire the knowledge is therefore extremely important. Appendix A reviews a number of the most commonly used techniques and tools.

2.5 A brief summary of the difficulties inherent in knowledge acquisition

For people who wish to generate software to mimic the human thought processes an understanding of the physical structure of the human mind, the organisation that exists within this structure, the multitude of different types of knowledge that it contains and how they relate, must be made clear.

As the literature demonstrates, this has not yet been achieved. Some headway has certainly been made on the structure of human memory with regards to the interactions of long-term and short-term memory and the limitations of both. However, investigations are still being carried out into the organisation of knowledge within long-term memory and how the different types of knowledge relate causing people act as they do.

It has been shown that different types of knowledge exist and that the state of a unit of knowledge can be transformed from the conscious to the sub-conscious. In addition, the associations between pieces of knowledge can be changed during the reorganisation that occurs following new/similar experiences, knowledge or cultural influences. These changes can lead to inaccuracies and inconsistencies in the knowledge held by an individual. Furthermore, the problem-solving and decision-making processes that are observed are fraught with that individual's beliefs, biases and value weightings.

Several techniques have evolved from fields such as psychology to assist a knowledge engineer in the extraction of some of these different types of knowledge. Gammack and Young (1984) and Welbank (1987) specified which techniques were appropriate for acquiring particular types of knowledge.

The debate continues, concerning the best psychological method, even though it is generally agreed that there is no one technique that can adequately acquire all the types of knowledge a human expert may use during problem-solving. Cordingley (1989) suggests that as many techniques as seems appropriate should be used. However, there has been no indication as to how the information should be gathered, using the various distinct methods, how it should be combined in a coherent manner nor as to how the different forms of the acquired information should be validated.

Machine-aided knowledge acquisition emerged as an attempt to speed up the process and, in some cases, to remove the errors and problems associated with involving a person not acquainted with the domain, i.e. a knowledge engineer. However, the field under review must already be formalised to enable the acquired knowledge to fit into a pre-defined model of the environment.

This is also true of the current shells and toolkits. For although they vary a great deal in features, flexibility and cost, the knowledge representation framework that is offered within these tools may well be inappropriate for the knowledge, the expertise and the problem-solving of the domain.

It is extremely difficult to determine the most suitable tool for a particular domain. It is well nigh impossible to succeed at this task if the domain, the domain knowledge, and the problem tasks to be tackled have not been examined first. Since, due to their constraints of pre-supposing the required problem-solving method and the structure of the knowledge base, i.e. the knowledge representation, these tools will in fact fail if they are confronted with situations outside their assumed boundaries. This implies therefore that a detailed analysis of the domain needs to be undertaken by a person who is acquianted with and understands the strengths and weaknesses of all of the available tools. Consequently, this person is highly unlikely to be the domain expert.

Here is yet another fundamental problem in knowledge engineering - the current mismatch between the representation of the knowledge expressed by experts and the form of knowledge required to drive a computer program, i.e. as data structures and algorithms. Experts still face the difficulty of trying to formulate their knowledge into a suitable representation for an expert system. This results from the inability of any available technique to adequately model the necessary knowledge for an expert system in a manner which is also natural for the human expert (Gruber, 1991).

It is the whole make-up of a person's memory that determines how a person acts or reacts: the hierarchical structure and how it is maintained, the types of knowledge and how they are organised, the prior experiences of the person, the biases, beliefs and values attached to items of information. As yet computer technology, even in its most advanced state, cannot possibly hope to emulate all these aspects of the human long-term memory. Therefore, since it has been shown that the structure employed for the memory must also be the same as that employed for processing (Schank, 1990), how can a computer be expected to behave and act in the same manner as a human during decision making and problem solving processes? Yet, this is the aim of expert systems.

Therefore it is still the incredibly time consuming and expensive process of knowledge acquisition and engineering that is the stumbling block for expert system development. As everything in these systems revolves around the knowledge obtained, the success of this initial stage influences greatly the performance, reliability, validity and usefulness of the end-system. To rely so heavily on one stage, the first, with no prospect of any benefits being realised until it is completed to a high degree of accuracy, seems hard to justify.

Consequently, instead of trying to mimic the human in such processes, as expert systems attempt and the AI field strives to achieve, a better goal would surely be to try to use computers to assist humans in those aspects of information processing where humans are weak and computers are strong. As Rettig (1993) states, 'cooperative problem solving systems serve as cognitive amplifiers of the human. Strong artificial intelligence is not necessary for a really intelligent solution. Instead a team made up of a person's natural intelligence enhanced by good computer software may be cheaper, more effective and more fulfilling to use.'

Expertise has been identified as requiring a large knowledge base with an efficient method of organising and categorising associated information and experience, thus enabling an expert to recognise and link previous similar events to the current situation under review. People are good at pattern recognition but they are poor at remembering all the previous instances of a particular situation and being able to quickly and accurately analyse similarities and differences between cases.

The indications are therefore that experts actually need tools that will enable them to investigate the strategies they use or have used, not the general statements but the precise strategies, and how these are applied to particular cases. If these tools are made available and are applied on the facts gathered then the biases of individuals should be minimised. The deviations in practice and in the interpretation of results will be laid open for the comments and the reviews of other experts, thus enabling focused discussions to take place with further investigations and experiments being initiated. Ultimately, for certain unformalised or ill-defined fields, the outcome will hopefully be a formalisation and standardisation of the domain knowledge and practices.

Consequently, developers need to build tools to assist the experts in investigating their fields in more detail enabling them to uncover explicitly the associations between facts. In other words, the domain experts interact with systems that are relevant to their tasks and which can produce benefits immediately, but which also permit data analysis to be undertaken to reveal links between elements and the characteristics of particular conditions. In this manner, a better understanding of the domain results, leading to a more formalised field with agreed standard practices and procedures. Once this has been achieved then a move towards modelling the domain can proceed, if it was deemed useful.

The following section gives details of the current research project which is investigating the feasibilty of this approach.

2.6 The current research

The characteristics and structure of the human memory, expertise and the different types of knowledge have been briefly outlined. Their effect on the knowledge acquisition process has also been described. It can clearly be seen why the knowledge engineering stage has been referred to so often as the 'bottleneck' in the development of expert systems (Waterman, 1986; Gaines, 1987; Hoffman, 1987). Yet the requirement for decision aids is growing, especially in medicine. Therefore when faced with developing a system to provide assistance in an ill-defined, unformalised specialist field, how and what computer system should be built?

In the field of knee ligament injuries, for example, the domain experts and the literature have indicated that there is a lack of agreement over the best strategy or treatment path to follow when a patient presents an unstable knee. In fact, individual experts themselves are not confident of defining the benefits of one treatment over another. Moreover, though numerous procedures exist for gathering information, there are no indications as to the value of each of the tests or which should be used during the normal assessment process. Therefore, experts simply select those tests they wish to use, determine the order in which the tests are to be carried out, and then they decide which treatment strategy they prefer to undertake at that time.

Futhermore, knee ligament injuries, as well as other medical specialities, have no set definition for 'success', mainly because patient expectation and circumstances differ. In addition, there are other questions, such as:

- how should the quantitative measurements and the subjective patient assessments be combined and weighted;
- is the information that has been gathered from various sources and/or by using different techniques of any actual value?

Consequently, in these scenarios acquiring experts' knowledge and modelling it in an appropriate manner are not the immediate problems. First, analyses must be carried out on the data and information gathered in the domain. Ideally, these analyses should occur at various specialist sites to enable examination of the various tests and procedures that are used by the different experts.

Therefore, rather than one pre-defined computer system being built and distributed, tools will have to be developed to permit appropriate systems to be constructed for the environment in which the system is to work, e.g. to ensure that the system will gather the data in the order defined by the specific expert and in response to the particular tests that the expert undertakes.

This requires that the tools be operated by the domain expert, who is highly likely to be a naive computer user. Otherwise, it would be extremely expensive and impractical to employ a computer programmer to build individual systems for the various institutions or specialist groups, who could be scattered world-wide.

Consequently, the aim of this research is to develop a methodology, which when it is computerised, will enable naive computer users, who are experts within specialist fields, to build their own Intelligent Database Decision Aid (IDDA). This IDDA end-system will store the domain data that is gathered, permit users to review these details and allow users to specify the data items to be analysed. Therefore users can take full control of the investigation being undertaken within their domain of expertise.

2.6.1 Scientific investigations

Currently scientific investigations follow the empiricist school of thought, i.e. conclusions are drawn from experimentation (see figure 2.4).

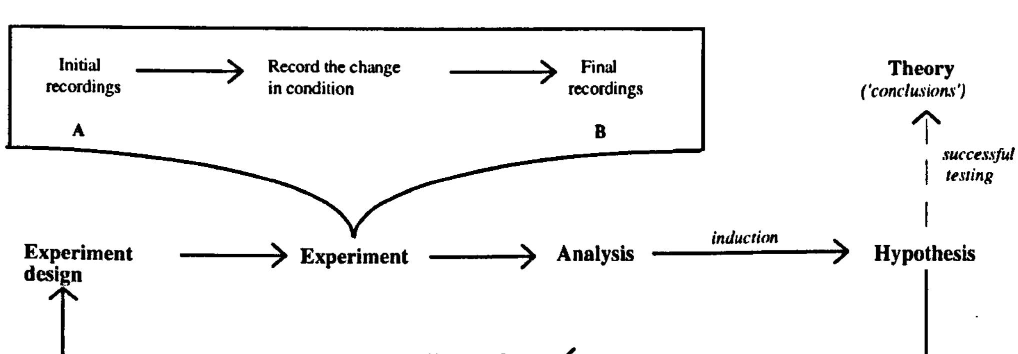


Figure 2.4: Empirical Investigation

The tests and procedures in B are a subset of A since A needs to also contain extra details of the person/object itself. To carry out valid statistical tests, data must be collected prior to an event and the same methods must be followed to acquire the corresponding data after the event. The differences between the two recordings can then be examined and conclusions drawn.

Specialist medical fields use this process extensively, resulting in the accumulation of patient case-histories. Generally, the filing and analysis of the data is still carried out by hand. This is labourious, slow, error-prone and thus expensive. It also restricts the number of reviews and analyses that will/can be undertaken on the gathered data. Consequently, some of the benefits which could have been gained from the study are lost. However, by computerising this process, the complexity and number of analyses can be increased at the same time as reducing the tedium and the number of calculation errors. Hence, correlations, associations and interactions between the data can be examined, leading to further insights into various aspects of the domain being uncovered and established explicitly.

It is very important that the same methods are used to collect the data and information for each case-history. Questionnaires have proven in the past to be good at prompting the investigators to always gather all the details and also at reminding them of the appropriate method to use. Another benefit of assessment via questionnaires is the freedom given to experts to work within their own guidelines and environments. Though it is true that experts within the same field may fail to agree on precisely which test to do, or exactly when in a particular assessment to do it, they do generally agree over the actual assessment in which the test should be undertaken, e.g. before, during or after the action. Therefore the only two restrictions in using a questionnaire are that: firstly, all the data, agreed by the consultants as being necessary, is gathered, and secondly, that no test affects the subject or item under review. Hence, computerising the questionnaires would enable the flexiblity to be maintained, thereby increasing the likelihood that a system will be accepted by the users rather than a system which dictates to the users precisely when and what should be carried out. A flexible approach is highly beneficial, for as Coles (1973) observed, 'it is relatively unusual in medicine for two places to have identical methods of operation and although the objectives of the system in the two places may be identical, local variations in day-to-day procedures may make a system, which works well in one place, clumsy or even unusable in another'.

Furthermore, if initially there is no agreement concerning which procedures or tests should be used during the assesment stages, statistical analyses on the data collected using the questionnaires can indicate which of the tests are of actual value and which are not. This could ultimately lead to a standardisation of the tests and assessment questions used within the specialist field. These would be firmly based on the gathered statistical evidence and not purely a result of the preferences and unsupported beliefs of each particular expert. For example, with the LRI knee ligament system, a number of the procedures within the domain are already coming under review. A current investigation being carried out using the LRI system is attempting to determine whether in fact X-rays have any diagnostic value in knee ligament injuries. If as suspected they are not, there are obvious benefits in the saving of vital time and money. Consequently, rules and procedures can be established for the specific specialist domain and the evolution of a better defined, more formalised field, with an agreed language and standard techniques, can begin.

Finally, if pre-defined questionnaires have been used to collect patient case-histories, statistical analyses and reviews could be undertaken to explain why a particular result occurred, e.g. what influenced one patient to responded to a particular treatment when another did not. Thus, when subsequent patients present similar profiles, more informed and hence better treatment decisions can be made. This should result in fewer re-admissions and more satisfied patients, thereby ensuring all the obvious benefits linked to such an outcome.

2.6.2 Computerising this process

To be able to obtain the benefits of computerising the scientific investigation process, individual end-systems must be constructed for each study. This is necessary to ensure that:

- a system fits the environment in which it is to work, i.e. that the appropriate domain langauage is used,
- data or information is only requested for the tasks actually carried out by the specific expert,
- these requests are in the correct order for the normal working practices of the expert and unit,
- the aims and objectives of the study are met.

Accordingly, the knowledge, expertise and experience of the domain expert are required to ensure that the end-system is relevant for the planned study. However, it is highly likely that the domain expert is not a computer programmer with enough experience to construct an appropriate system from scratch. Hence, a suite of computer tools is required that could lead a naive computer user through the development task. Therefore, these tools will need to encapsulate the pertinent skills of a computer scientist to enable them to automatically construct an appropriate system for the study from the details supplied by the domain expert.

Consequently, this research intends that domain experts will be 'builders' of their end-systems, thus they will closely identify with their systems. There will be no knowledge engineers or computer programmers involved. Domain experts will be responsible for defining the facts to be collected and stored in the database, for specifying the questions and options that will be available to obtain these facts and when these questions are to be asked, i.e. they define the assessment questionnaires. The IDDA end-systems will then be automatically constructed from these specifications. Therefore experts will 'understand' their IDDA systems, i.e. no question will be presented or asked without them knowing why. Finally, experts will also have complete control over both the selection of the facts/data to be reviewed/analysed and the choice of technique to be used. Domain experts will thus dictate to the systems and not the systems to the experts. Hence, the IDDA end-system enables the domain expert to analyse, theorise and investigate hypotheses. The ultimate aim is to provide end-users with facilities to assist them in their daily activities and to aid them in making more informed and hence better decisions.

By using the approach of domain experts building their own system and of basing the constructed end-system on the empirical method, many of the knowledge acquisition pitfalls previously encountered by the expert system approach can be avoided:

a) it is not necessary for experts to articulate their underlying knowledge and expertise, just to be able to specify the facts or data that are important in their specialist fields, i.e. the signs or symptoms that require to be collected,

b) domain experts should feel less threatened since their problem-solving techniques will not be laid open for examination or questioning,

c) computers will be viewed purely as just another tool to assist experts in the areas where humans are weak, e.g. in the collection, review and analysis of large amounts of data and information, rather than as a replacement mechanism to undermine an expert.

The question is therefore whether or not domain experts are capable of devising/revising an appropriate set of standard questionnaires for their specialist fields that will capture all of the data required for their studies. To undertake any clinical investigation, whether it is conducted manually or via computers, domain experts need to be able to determine which facts (measurements, signs, symptoms) must be collected and when.

As the literature previously showed, experts have a large base of knowledge gathered over time and experiences. They possess the ability to identify relationships which exist in this information and use them to answer questions and develop theories. They can adjust to the level of a novice during explanations and select justifications for problems that a novice could follow. This all demonstrates their ability to break problems down into manageable portions.

Hence, experts are neither required to specify or to justify the processes by which they select knowledge nor to describe precisely the procedures used to solve a problem or to reach a decision (which are the major stumbling blocks in the knowledge acquisition process). These types of knowledge do not need to be explicitly stated. Rather it is envisaged that the techniques already used by experts, when they question and examine their own domain, will be utilised. Consequently, when domain experts undertake an investigation in their own specialist fields, they know the information that requires analysis and the tests/trials needing to be carried out to obtain this data. Therefore, they know the general information that must be gathered, when and how. Consequently, domain experts should be able to produce the standard structured questionnaires required and describe the procedures that should be followed (see figure 2.5).

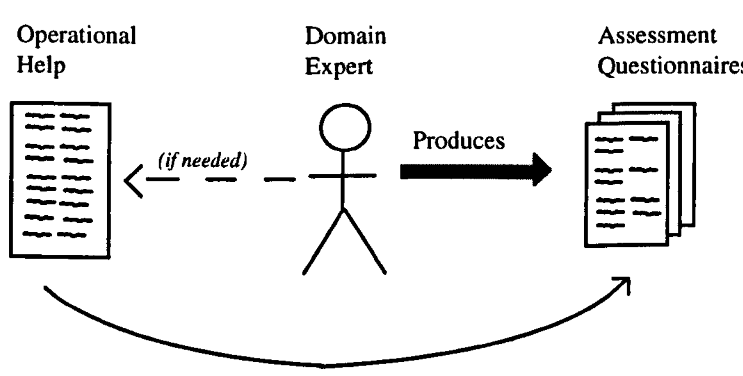


Figure 2.5: The domain expert's contribution

A number of the benefits of this approach have already been outlined:

a) the avoidance of the knowledge acquisition difficulties encountered during expert system development,

b) the advantage of having the full involvement of the intended end-users in the project, with respect to user acceptance and commitment,

c) the construction of a system that can assist in the daily operations of the unit,

d) the reduction in time, effort, and monetary costs in constructing an end-system,

e) the adoption of standard questionnaires and procedures, thus aiding novices in the collection of complete records (de Dombal, 1984),

f) the ability to operate in unformalised or ill-defined domains.

However, this approach is not totally trouble free. Experts have little spare time to either operate the end product or to learn complicated instructions. They must feel 'comfortable' with the system and not feel threatened, or be made to look stupid either by their inability to operate a computer, or by their inability to understand or follow the results produced by a computer. The commitment and support of the envisaged end-users are essential for the success of any computer project. Therefore, the whole environment (that of the computer tools and the generated IDDA end-system) has to be designed and developed for naive computer users.

2.6.3 Task allocation

Domain experts have little spare time. Therefore there is a need to keep their involvement in the construction phase to a minimum. They know the domain in detail, the working practices of the unit, the aims and objectives of the study about to be undertaken, the details requiring to be collected to achieve these goals and the procedures that will acquire these necessary details. They can also explain and describe any of the more complex or unusual operations to help less experienced personnel obtain the required data in the correct manner. Their input to the development process should therefore be concerned with this kind of expert knowledge and not the rudimentaries of operating the computer. However, they will be deeply involved with the analysis and comparisons of the stored data when they investigate their hypotheses and theories at a later date. Nevertheless, for this initial stage, they will have little or no direct interaction with the computer tools. Their contribution, therefore, to the building of the end-system, will be restricted to:

- specifying the information to be collected for those investigations by defining the appropriate questionnaires for each assessment stage,
- defining any help text that they perceive as being necessary to ensure that an investigator will carry out the assessment in an appropriate manner.

In the case of a medical system, junior doctors will generally be given the major task of interacting with the computer tools during the construction phase. For example, they will be required to specify the type of answer that the computer should expect as a response to a particular question in an assessment questionnaire, and to link any defined help to the appropriate assessment questions. Furthermore, it is highly likely that they will be the people responsible for entering the patient details and information gained during the patient assessments once the IDDA end-system has been built. However, even though their time is not as crucial as the experts, they are not expected to waste it in attempting to type into the computer the actual questionnaires, which usually involve substantial amounts of text.

Secretaries have the necessary keyboard skills. Moreover, they will generally have had the invaluable experience of using a word-processor. These skills should be utilised to construct appropriate files, for the computer tools to access, from the bulky assessment questionnaires and any instructions/help that the doctors believe may be required to complete these forms.

With this technique of utilising the different abilities (see figure 2.6), skills and knowledge of the people involved in the specialist field, the tedium of having to complete tasks for which individuals are unsuited, and consequently find difficult, can be alleviated. Usually the more experienced the users are in a particular skill, the less time it takes them to accomplish a task. As the following chapters will show, naive computer users become quickly disillusioned if, to accommodate a computer, they themselves have to adapt drastically not only their knowledge of the domain but also the skills within the domain that they have previously acquired. Furthermore, they are loath to use systems from which it is hard to see the results of their efforts.

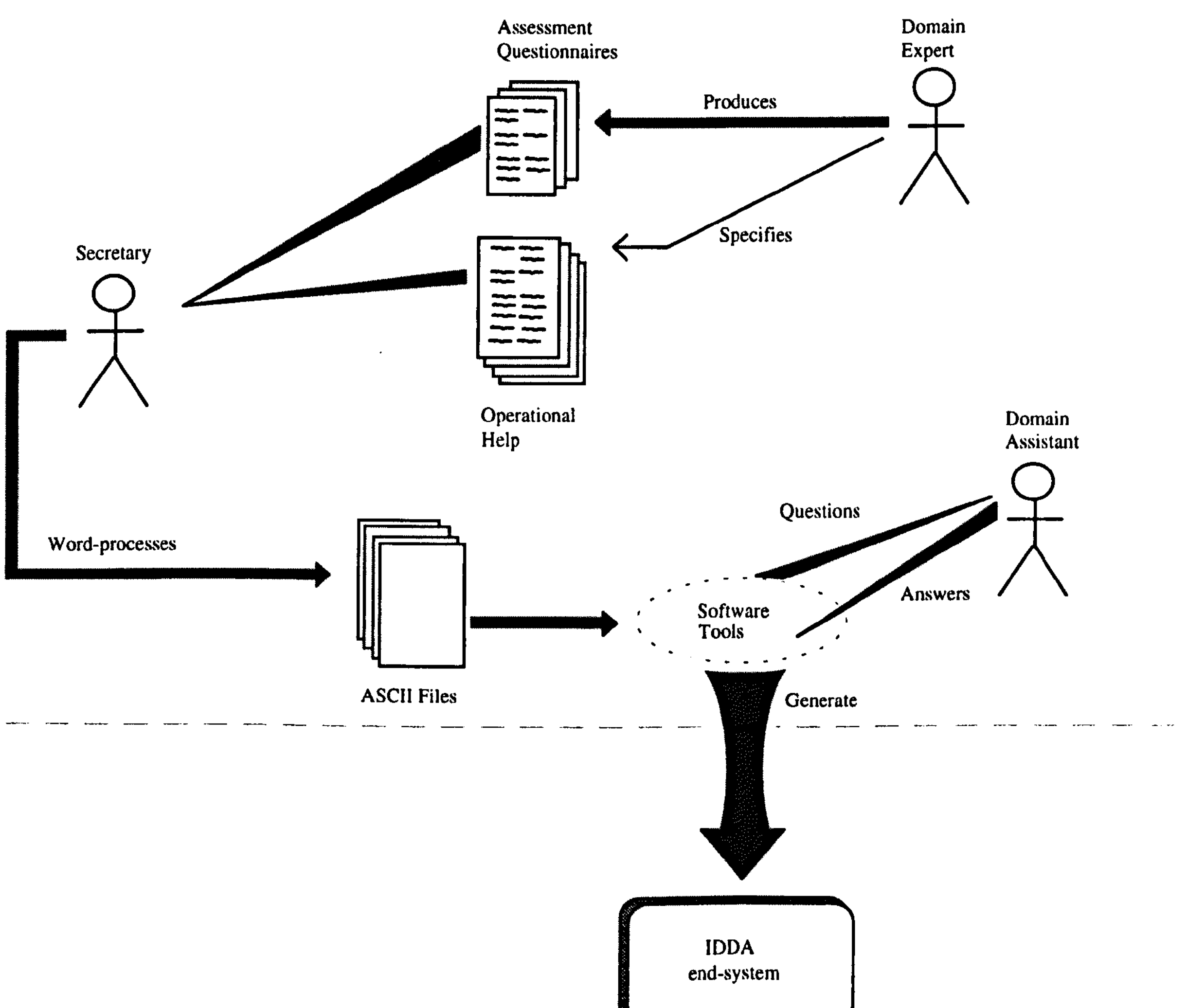


Figure 2.6: Task sharing within the specialist domain

Consequently, the sources of knowledge and skill from the specialist field that will be utilised during this project will be :

a) the typing and word-processing skills of the secretary,

b) the general knowledge of the working practices and procedures of the domain assistant, i.e. a junior doctor, in the case of a medical system,

c) the detailed specialist knowledge of the domain expert.

The next chapter reviews various aspects of system design from life-cycles through user participation in the design process to whether the development is viable and the types of user interfaces and interaction techniques.

Chapter 3

Aspects of System Design

3.1 Introduction

This chapter provides an overview of a number of the most important aspects involved in designing and building computer-based systems. The issues considered formed the basis of the major decisions taken during this research project, but a description of the actual development process is discussed in more detail in Chapter 4.

The cost of software for a computer-system does not stop once the software system has been developed. For example, a system requires maintainence after it has been delivered. Whilst hardware costs have dropped dramatically during the past decade, software costs have escalated, becoming the largest expense in many computer-systems. It was discovered that as systems grew larger, their quality became more suspect and control over projects diminished. To alleviate these problems, a set of techniques evolved. Collectively they became known as the 'software engineering' techniques. These deal with software as an engineering product requiring planning, analysis, design, implementation, testing and maintenance (see Alonso et al (1990)).

The following sections review the design process from the software life-cycle, through user participation in the design process and the product viability questions, to the crucial aspects of the user interface, user interaction and the impact on the users' work and environment.

3.2 Software life-cycle

The software 'life-cycle' is concerned with the development of software from initial concepts through to delivery, use and maintenance. Generally, it is not possible to proceed directly from the initial concepts to executable software, instead the problem is divided into manageable parts. This enables a designer to deal with smaller units which are easier to construct, verify and modify. These steps frequently produce more detailed descriptions outlining the proposed system, for example, requirement specifications and design descriptions (McDermid, 1987; Boehm-Davis and Ross, 1992).

The first software development life-cycle that had an impact on conventional software development was the 'waterfall' life-cycle (Wilson et al, 1989). This is illustrated schematically in figure 3.1.

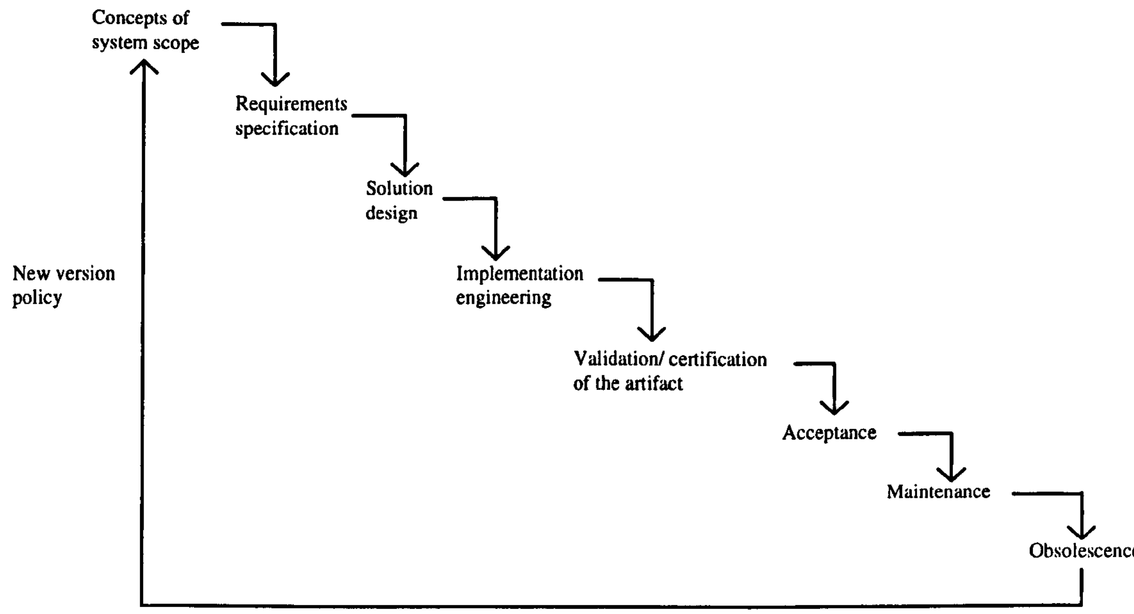


Figure 3.1: The classical waterfall software life cycle (in Wilson et al, 1989)

The development process proceeds sequentially through the identified phases, when one finishes the next starts. In this traditional life-cycle, communication occurs between the developers and the users at an early stage (e.g. the 'Requirements specification' in Figure 3.1) to determine the information and system requirements. After these initial sessions, the developers tend to work in isolation until they deliver the system. Therefore this model relies on the assumption that the requirements are stable and frozen. However, in reality this is rarely the case.

Users can verify the proposed system during these initial sessions but only by reviewing a mass of printed documents. Consequently, the users find it 'difficult to judge exactly whether the proposed system satisfies the original requirements, because there is no tangible demonstration of the system functions' (Lea and Chung, 1990). Thus, users need to gain some experience of using the proposed system before the system requirements can be accurately and adequately specified. Hence, with this traditional approach, these changes in requirements only become apparent after the initial implementation, often as a result of users claiming they are dissatisfied with the system and that they find it hard to learn and use (Carey, 1990).

This method, therefore, prevents easy, early modification of any work. Moreover, the importance of undertaking careful requirements analysis has still not been fully recognised, resulting in systems being designed which do not adequately support those activities which motivated them (Byrd et al, 1992).

In addition, as a system progresses through the life-cycle and becomes more complex and complete, changes at later stages are more costly in terms of effort and resources than those at early stages. This is because the impact of the changes become greater the further through the cycle they are made. Estimates have been made stating that the removal of an error at the requirements analysis stage costs, at least, 100 times less than when the system is in operation (Diaper, 1989; Byrd et al, 1992).

The realisation of the high cost of changes in a system's design, prompted a move towards a more dynamic process. This involved illustrating the current stage to the intended users, as well as undertaking exploratory work into later stages, and thereby obtaining essential feedback with which to direct the development process.

Prototyping refers to developing quick, computerised solutions to problems. It is built around the assumption that the system's requirements and goals will change during the development cycle. Thus, a developer demonstrates the prototype to the users and obtains feedback regarding their views. These suggestions then drive the development of the next version of the prototype.

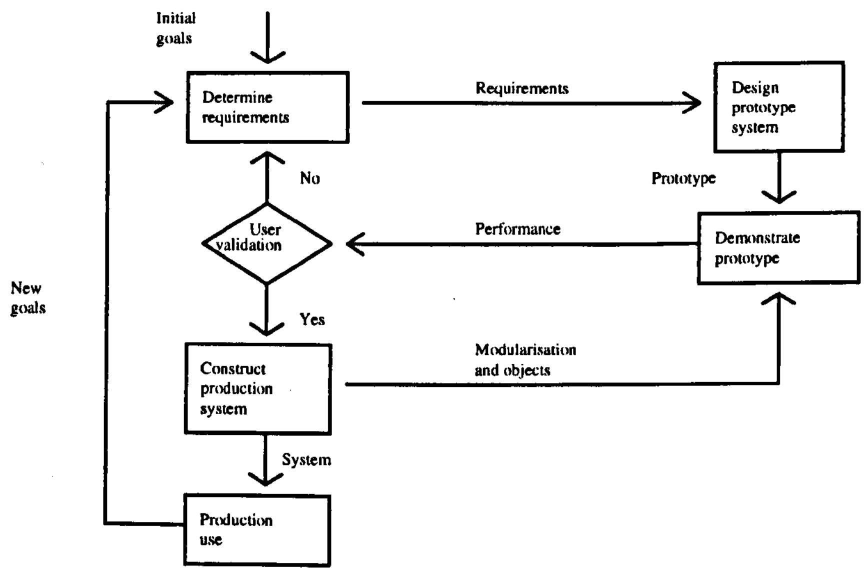


Figure 3.2: The prototype cycle (in Luqi, 1989)

Consequently, valuable time and effort are not invested in implementing inappropriate facilities and functions into a system. This is particularly useful when the developer is unfamiliar with the application area or is uncertain over the requirements (Lea and Chung, 1990; Roberts, 1990; Luqi, 1989). Prototyping therefore enables the review of the stage specifications as the system is evolving, which aids the discussions and finalisation of system requirements and helps to determine whether the requirements specified have actually been met.

There are basically three different kinds of prototype (Tate, 1990; Jojo, 1994):

- rapid - prototyping focuses on essential aspects of the problems at hand and minimises effort spent on other aspects. With rapid prototyping, quick methods of construction are used. The developer then discards earlier versions of the prototype as it becomes obsolete and begins afresh. It has been successful during the development of small systems (Roberts, 1990),
- incremental - the system is broken down into parts, each of which is prototyped and included one by one into the finished system,
- evolutionary - the prototype is refined and refined until it becomes the real system by satisfying the user's requirements. In fact, maintenance is viewed as a mechanism for ensuring that the system is kept up-to-date, hence the system development never really finishes.

Carey (1990) lists a number of the advantages of prototyping:

- systems can be developed more quickly,
- systems are easier for end-users to learn and use,
- prototyping facilitates end-user involvement. As Duchessi and O'Keefe (1992) explain, 'developers create an environment that gives the users a sense of real participation which increases the chance of success',
- system implementation is easier because users know what to expect, 'the prototype should be perceived and experienced by its users in as realistic a way as possible' (Jojo, 1994),
- user requirements are easier to determine. Berry (1994) agrees, 'users seldom know what they want from a system. It is difficult for them to imagine what new technology can do for them',
- development costs are reduced,
- the final system is 'correct' and is accepted by the users. Since this approach enables user preferences and ideas to be easily incorporated into the system design, it aids user acceptance of the finished system.

However, there are drawbacks:

- unrealistic user expectations - especially if the users perceive many of their early suggestions and wishes being incorporated,
- inconsistencies between prototype and final system,
- final system inefficiencies - 'maintenance and modifications to expert systems and their knowledge-bases, especially those produced through rapid prototyping, is incredibly expensive and may well be nigh impossible' (Diaper, 1989),
- little formalised structure, unless it is imposed rigorously on the developer, i.e. by embedding better project planning, project management, validation, etc. As Jojo (1994) points out, 'with any development methodology, a budget, resource and expected milestones and deliverables have to be 'mapped' onto the various phases. Proponents of prototyping have typically given scant regard to these aspects of development but in many organisations it is not possible to embark on any project without some budgeting and allocation of resources'.

Consequently, developers began to believe that neither prototyping nor structured techniques alone could solve all the problems encountered during the various development stages of computer systems (Ratcliff, 1987; Yourdan, 1986).

Whiteside et al (1987) and Carroll and Rosson (1985) both outline methods by which usability specifications can be used as checklists and guidelines in an iterative design cycle. They believe that these objectives can provide precise and testable statements of performance which can be reviewed and evaluated throughout a design process. These goals therefore help to focus attention on the critical components of the proposed system and provide a method to aid resource management.

Weitzel and Kerschberg (1989) combined various approaches in the development life-cycle of MEDCLAIM, a knowledge-based system for dealing with medical insurance claims. A prototyping methodology using expert system shells and programming environments was used initially. The development path was a series of 'processes' that were activated, deactivated and reactivated as the system evolved. Weitzel and Kerschberg describe in detail each of the processes and the tasks required at each stage. The prototype was then used as the specification for the development of a fully operational system using conventional system development life-cycle methodologies. They believe this combination of methods would 'strike a balance between unbridled flexibility and stultifying structure'. The benefits of both approaches could thus be gained, whilst avoiding some of their limitations.

A variety of design tools and construction kits do exist to aid developers (Fischer and Leuke, 1988). Yazdani (1989a) discusses the pros and cons of shells and toolkits. He describes programming languages as 'building blocks', toolkits as 'prefabricated houses' and shells as 'prebuilt houses'. These descriptions emphasise the flexibility and the possibilities of each of the different techniques. Versatility, however, can create its own problems:

a) in the difficulties in learning the programming language,
b) in the difficulties in learning the design strategies and the characteristics required to develop an system,
c) in the cost of purchasing such software, e.g. shells and toolkits,
d) also, in the case of toolkits, the amount of computer power and resources required to provide the facilities offered.

The end result is often a choice between system efficiency and high development cost, both in terms of human development and system development. As Thomas (1983) explains, even if it was discovered that people could dial telephones more quickly and accurately by using an eye-tracking camera rather than push-buttons, it would be unlikely that such phones (costing 50 times as much) would be widely used.

With this research project, the decision concerning the software development cycle was determined, in part, by monetary costs. Any system developed had to be cheap, both in terms of hardware and software. Therefore, this automatically restricted the development platform to a PC with programming languages being used for the construction of the end-system. Since software tools were to be built, the use of programming languages would give the extra flexibility required during the development of the tools. The tools, themselves, could then be constructed to give all the required support to a developer of an IDDA end-system.

One major benefit to this project was the existence of a fully operational end-system already in use within a specialist domain at the LRI. The system, as explained previously, had been designed and built as a one-off solution to assist orthopaedic consultants in assessing knee ligament injuries. However, the success of that initial system, prompted other enquiries for similar applications for different specialisms, thereby initiating the current research study.

Consequently, this initial knee ligament system could be reviewed with the domain users as though it was a first stage prototype. The users, who were orthopaedic surgeons, were very happy with the system. They said that they had found it both easy to use, though in some parts rather long-winded, and easy to learn, which is one of the major concerns when building such medical systems. They did however request additional facilities for evaluating the data collected and agreed that extra features to assist in faster data entry would be beneficial. Nevertheless, they could act as a user group, making comments and reviewing subsequent prototypes. In addition, the domain could be used as a test-bed for this research study.

There was however one problem. This LRI system was in daily operation and so changes to the data entry programs could cause confusion, time delays, and effect its robustness. It was therefore decided that, since the end-users could explain quite clearly, with the current system, their views regarding the data entry screens and the extra facilities that they desired, the basic LRI database system would not be altered at this stage. Instead, the users would wait and then comment on the first end-system developed by the tools, which would include these features.

With regards to the requests for additional facilities to assist in the data analysis, it was agreed that these would be built separately from the original LRI system and then be integrated in once the end-users were happy with them. Consequently, the LRI end-system progressed using an incremental prototyping technique. However, it could be viewed as using an evolutionary approach as well since the data entry facilities and screens were to be changed and up-dated at a later stage.

For the development of the tools, a combined approach was used (see figure 3.3). First, a structured technique was used to review the LRI system to determine the requirements and specifications of the generic systems that the tools were to produce (e.g. requirements specification). Having established these, an investigation was undertaken to ascertain the domain knowledge required to build such end-systems and to decide upon the best method available to acquire this knowledge (e.g. knowledge acquisition, see Chapter 2). The stages involved in constructing an end-system were then determined and an evolutionary, incremental prototyping approach was used to design, implement and test (in-house, one of which was the original LRI system) these stages. Evaluations were then carried out by running trials. These involved computer competent but domain naive users, computer naive and domain naive users, and, computer naive but domain competent users (see Chapter 6 for more details).

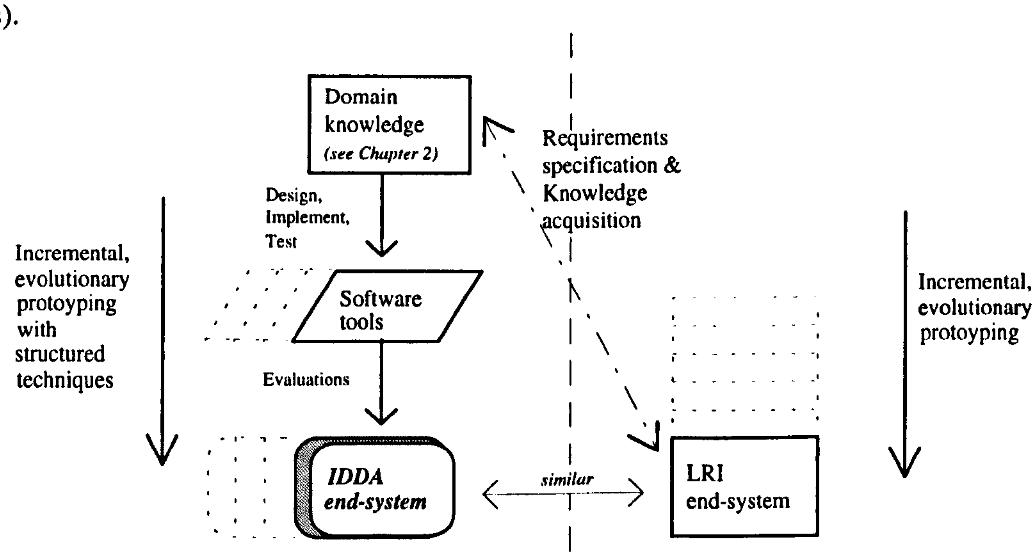


Figure 3.3: Development adopted for production of the tools and the IDDA end-system

One of the fundamental reasons for emphasising the prototyping approach was the belief that end-users should play an active part in developing their system, since they will ultimately be the operators. When constructing software tools, which could be used in a number of domains, it is harder to ensure that there is some end-user input. However with this research project there was the major advantage of already having a similar end-system in daily operation. The importance of user involvement in system development is reviewed in the following section.

3.3 User participation in the design process

One of the major problems designers often encounter in developing software is that users continually 'change their minds' and the specifications during system development. This, therefore, adds further weight to the argument that a prototyping approach best suits such situations since users are able to reconsider their requirements frequently. Thus any alterations to the specifications of each prototype will be minor in relation to the same changes being requested at the end of the development process.

Experience has shown that without a clear understanding of user requirements, system developers fail to provide crucial facilities which results in the production of end-systems with limited utility (Berry and Broadbent, 1987). Not only are the needs of users important during system development, but the properties of the domain and the task requirements must also be analysed and reviewed. Generally, users are asked to comment on a system in formal or informal evaluation sessions after a system has been built. However, this is often too late. Discussions with users and the involvement of users throughout the development process should occur from an early stage. This will ensure that a better understanding of both the constraints imposed on the system, and its required capabilities, have been appropriately interpreted and implemented by the developer (Parsaye and Chignell, 1993; Berry, 1994).

Furthermore, the potential users must perceive a need for the system. As Keyes (1989) correctly points out, 'without user support, systems won't work'. This was soon apparent to an information systems department in a U.S. Post Office when it tried to automate the work of the clerks who served at the windows. The attempt was met with an enormous amount of resistance that even stretched to sabotage. This delayed the implementation for years. The system was a conventional computer application but it lacked the crucial user involvement and support for the project. 'The users must perceive a need. Then they must see the benefit because we can't force them to use the system' (Keyes, 1989). Roberts (1990) agrees, stating that this obvious aspect of system design is, at times, ignored, even though 'the chances of producing a successful system are greatly reduced if potential users do not want it'. One method by which to achieve this, and thereby greatly enhance user acceptance, is by allowing users to have an active role in influencing how the system will look and work.

In fact, as Jojo (1994) discovered, the involvement of a user group in development discussions could reveal both operational and security concerns that had previously been overlooked. Moreover, by involving the users in simulation sessions, common user problems and errors can be identified and resolved prior to the system being installed, provided of course that the simulation scenarios were realistic (Booth, 1991; Berry, 1994).

DEC discovered the benefit of user participation during the development of their XSEL system. XSEL was designed to aid the sales force in configuring systems accurately and quickly. From the start it was decided that users should be the dominant partners in the development team. Consequently, a user design group was established. When XSEL became operational, the sales force proved to be fully committed towards its success. They understood the system, as the application had been designed to meet their needs, and they were prepared to spend time developing it further. It also proved to be a major financial success to the company. Considering the success of this project Mumford (1988) asks, 'participation works well, it produces results. Why [then] is it still rarely used in any significant way?'

Mumford (1991) believes one possible reason is that many people do not know how to organise such participation. Therefore, it is still common for the design of a system to be neither based on the information and knowledge required by the users, nor on the best method to present the information to support the users' tasks. Instead, system designers' decisions are based on their specific knowledge about the technical system, the available technology and their 'common sense' judgements about the information the users need (Sundstrom, 1993; Braunstein et al, 1991). Consequently basic factors such as the use of correct vocabulary, the removal of irrelevant questions, the correct presentation and sequencing of tasks and the ease of use, cannot be guaranteed (Fieschi; 1990; Todd and Benbasat, 1992; Berry and Hart, 1991).

Users are only interested in carrying out their required tasks quickly and with minimal trouble and effort. They do not want to have to spend time searching for a method to achieve their tasks. Consequently, designers must be aware of the target population, their tasks and appreciate any external constraints and influences that might effect user performance. Only if this occurs can the system be designed to be used to the greatest degree of efficiency (Wasson and Akselsen, 1992; Sundstrom, 1993; Oborne, 1987). Those functions and facilities which are developed because they are 'clever' rather than because they play a useful role in a typical user task, can easily confuse inexperienced users. As Berry and Broadbent (1987) point out, 'interacting with a system which has multiple windows popping up and disappearing all over the screen can be very disorientating.'

There is evidence to show that a key ingredient in a successful project is user participation (Berry, 1994; Byrd et al, 1992; Lings et al, 1991). Gordon et al (1987) therefore ask, 'why did designers ignore human factors principles for so many years, particularly when the stakes were often high?' To answer this, Gordon et al have proposed five reasons:

a) in complex systems, designers are immersed in a large number of design details (e.g. reliability, cost, operating conditions) and therefore cannot see the importance of, or do not have the time to deal with, human requirements,

b) many designers treat human factors as 'common-sense', not requiring specialised training. A common-sense approach may offer improvement but it rarely optimises the design,

c) designers of complex systems, out of necessity, become so engrossed in the design process that they often assume that a variety of 'facts' about a system are obvious. Unfortunately, a user only knows the system from the way it presents itself. It is often difficult for the designers to see the system from a user's perspective, and they therefore cannot always recognise that some control or display element will be obscure or difficult to use,

d) designers tend to use the intelligence, flexibility and trainability of the users as an excuse. If the final design does not produce the desired level of performance, a more detailed manual or more user training are used to solve the problem. This reliance on training or memory places an additional burden on a user. The problem is especially acute when users operate the system infrequently, or in situations of high mental stress, when users tend to revert back to old habits,

e) designers in all areas tend to take a lot of pride in their design. Consequently, they tend to be defensive in response to suggested alterations in their original design.

However, human factors are important issues and should not just become fashion accessories to be added afterwards or reviewed only if costs allow. A system has to operate around a user and not a user around the system. If the environment and pressures placed upon users are not considered then the acceptability of, and willingness to use, the system will disappear and any benefits that the system might have brought will be lost.

These considerations regarding human factors have to also include the actual working practices of users. The strategies devised in the design phase should closely follow those procedures currently undertaken in the manual system. Otherwise, by insisting on new or extended practices, there is the danger that the end-users may totally reject the new system - 'many physicians completed the form only if time allowed, and some stopped completing them altogether' (Innes et al, 1985). However, involving users throughout the design will help ensure that the system provides the necessary facilities for the required tasks and has the capabilities to perform efficiently, without any significant change in user operation (Nickerson, 1981).

With this research study, the decision to follow a predominately prototyping approach during both the continual development of the LRI system and the evolution of the software tools, enabled user participation to be integrated into the development processes. Hence, feedback from the users of the LRI knee ligament system could be incorporated into the design of the IDDA end-system constructed by the software tools. Thus, any change to the IDDA end-system also effected the design of the software tools themselves. In addition, testing and undertaking in-house trials with the tools provided information on their accuracy, ease of use, sequencing of tasks and on minimising the time and effort required to complete a phase in the construction of an IDDA system.

However, care had to be taken regarding the suggestions of the orthopaedic users to ensure that the eventual IDDA system was not merely designed for the one environment or task. Generic tools were being developed from a methodology which was attempting to provide a mechanism for naive computer users, who are domain experts, to develop their own IDDA end-system. These IDDA systems would assist experts to undertake empirical investigations within their domain. Hence the facilities and features within the end-system had to be appropriate for such a task and not solely within the knee ligament domain.

Nevertheless, that specialist orthopaedic field provided useful information regarding:

- the processes involved in undertaking empirical studies,
- the typical reviews and analysis carried out on gathered information,
- the constraints imposed by this type of dynamic environment, and,
- the common difficulties and fears experienced by naive computer users.

Hence, it provided a revealing insight of the likely users, tasks and environment that both the software tools and IDDA end-systems would have to confront and operate with. It also emphasised the importance of constructing an IDDA system that incorporated the current working practices of the domain during an empirical investigation. For, it is believed that, by modelling the current practices and thereby providing little if any overhead during the task of data collection, the benefits of computerised records for tasks such as information review, retrieval and analysis would be self-evident to any prospective user. Therefore, the justification for constructing an IDDA system using the software tools could easily be made by explaining the advantages that would be gained during these types of task.

However, such justification would rely, in part, on three factors: the time required to construct an IDDA system; the time required to operate such a system; and, the effect on the user's work and environment. These are discussed in the following sections.

3.4 Time required to develop a system

It is important that the users know the approximate length of time required to develop their system, especially if they are expected to participate in the development process. User commitment to the project will only be gained if the users perceive that the benefits of such a system outweigh the time and effort expended to produce it.

The development delay was particularly problematic for expert system construction, especially since by their nature they require substantial amounts of user participation. The production of a prototype typically took 6 to 24 months and full-scale systems were in the order of years (Duda and Shortliffe, 1983; Steels, 1987). Hoffman (1987), listed some examples : INTERNIST took 10 years with the help of a full-time specialist in internal medicine, and, XCON took 2 man-years with a team of a dozen researchers (and is still being refined).

Consequently, in an attempt to alleviate the problem, a number of suggestions for restricting the system were made:

a) carefully selecting the domain - As Davis (1989a) explains, "Is this an unstructured domain where the best performers can and do become intuitive ? Is this a domain where an expert has a gut feeling almost instantaneously about what's going on ?' If it is, stay away from that domain as far as making promises about capturing the expert's understanding and skill in the form of complex reasoning system using facts, rules and inferences. Medicine, prospecting for minerals, and all policy-level decision making, to name a few are domains where the best performers seem to develop and use intuition. These are also domains where expert systems, contrary to what you may have heard, do not measure up to experts',

b) ensuring that there is a clear problem specification - 'Too frequently the failure of the development team to precisely understand the project's goal dooms the project from the outset. An ill-defined problem produces at best no solutions and at worst bad solutions. Formulating an accurate statement of task is obviously more probable for a tightly restricted problem' (Bielawski and Lewand, 1991),

c) restricting the scope of the system - 'A rule of thumb frequently offered is to restrict the scope of a system to a task that, if performed by a human, would require no more than a few hours. Imagine trying to capture in a knowledge base the expertise of a person who takes weeks of months to work out a solution to a problem! The mental processes that the human problem solver invokes during this amount of time are overwhelming. Considering how much common sense the human must use, the idea of developing a system to address a problem such as this becomes bewildering' (Bielawski and Lewand, 1991),

d) ensuring that the consequence of failure is manageable - Davis (1989b) is willing to bet that every expert system built in the next 10 years will fail while in real use, 'consider XCON, with 80,000 examples run (as of August 1985) and still knowledge is incomplete. The errors are infrequent, but inevitable, as in any incompletely understood task'. Therefore, even with 80,000 test cases, errors can still occur. However, when the consequences of failure are manageable, as Davis (1989b) explains, with 'XCON, for instance, currently has roughly a 2% chance of failure but even then the consequence is inconvenience and perhaps a small monetary cost, things we can well afford'.

However, even with these reduced boundaries, the time taken to develop an operational expert system is still significant. Moreover, with an expert system, the system is not operational before a substantial portion of its knowledge-base has been constructed and verified. Hence, the intended users are not able to gain any benefit from such a system until it has nearly been 'finished', even though they are expected to contribute significant amounts of time and effort.

In addition, building the knowledge-base, itself, causes many problems, not just in knowledge acquisition and representation, but in accommodating new facts. As more and more facts are gathered, the list of rules grows. This in turn effects the speed of the system and more control processes require to be inserted. Such an approach ignores the problem of how human knowledge is integrated with other knowledge and reorganised over time (Kolodner, 1983). Therefore there is a need to refine, reorganise, rebuild and retest the knowledge-base at each stage of development. This iterative process is extremely time-consuming and laborious for both developers and users.

With more conventional systems, specialist knowledge is still required, though on a much smaller scale. It is used to ensure that the system is appropriate for the environment, the tasks and the users, both in terms of its actions and its interactions. User participation in the development process is therefore important whichever type of system is being constructed.

User commitment and backing will generally be enhanced if a system can be produced rapidly. If prospective users can see an end result occurring in the near future, they are more likely to agree to participate at the outset of a project. This end-result need not be the fully finished system, it could be an intermediate system that undertakes some of the tasks but not all. The other tasks could be developed as separate units and integrated at a later date, i.e. using an incremental prototyping approach. Hence, the users could obtain certain benefits from the evolving system near the outset and these would increase as the development progresses. Therefore there is an obvious correlation between the expended time and effort and the benefits the users gain. Furthermore, if only limited amounts of each person's time are required, an individual is more likely to be fully committed in those short periods and be more willing to allocate the necessary time to the task.

This is the approach adopted for the development of an IDDA system. The initial stages in the construction are divided between the different personnel within the unit, as described previously in Chapter 2. Consequently, this reduces the involvement of the various participants to only those tasks for which each person is best suited and therefore would find relatively easy.

This utilisation of these different skills, and the limitation of the amount of time and effort each individual would have to expend, during a system's development would therefore enhance user commitment to the project. Without user support, an IDDA end-system could not be built since the methodology envisages that the users will be the builders of their own end-systems.

The actual development time required for the construction of an IDDA system is also dependent upon the users, i.e. it is reliant on:

- the time taken by domain experts to finalise the questionnaires and produce any help text,
- the length of these questionnaires and help text which require to be word processed by the secretaries,
- the length of these questionnaires and help text which need to be defined by the domain assistant.

Individuals can work at their own pace and can determine the priority of the work with respect to their other daily tasks. However, once all the information has been entered, the tools can automatically construct the specified IDDA system in a matter of minutes. The entry of data for the empirical investigation can then begin. Therefore the users quickly realise the benefits of using such a system as the initial reviews and data analyses on the collected information.

Once the time taken to develop an end-system has been justified, the users' attention is usually directed towards the time required to operate the implemented system.

3.5 Speed of operating the implemented end-system

Producing systems that are not time critical, enables users to consider at leisure the program's response. Consequently, although it may be desirable to accumulate the domain knowledge and information to be able to undertake analyses, constructing systems to be used concurrently in real-time decision-making may not be required or be the best approach to be adopted, especially as they are highly likely to interfere with normal human decision-making processes (Davis, 1989b). Therefore, users require systems which assist them in achieving their goals but do not increase the work pressure, the stress, or the time involved in completing their daily tasks.

For example, doctors hold brief patient consultations under severe time pressures. In addition, they have varying clinical experiences and skills. The nature of their job requires them to make important decisions regarding the preservation of their patients' health. They do not want additional external activities interrupting or confusing them during these consultation periods. With MYCIN, the question and answer interaction was reported to have taken between 30 and 50 minutes to reach a reasonably accurate diagnosis and required the doctor to answer a string of relentless questions that seemed to be irrelevant to the situation (Berry and Broadbent, 1987). Consequently, it is important that computer systems give users the help and information they need at a time that is convenient to the user and not at some other pre-arranged time which seems optimal for the system (BCS Committee Interim Report, 1983).

In addition, when users operate computers, they want to accomplish tasks but do not want to invest time in learning and perfecting skills which enhance their effectiveness and efficiency in using the computer systems (Wasson and Akselsen, 1992) Santhanam (1993) termed this the 'production paradox', where the strong desire of individuals to achieve their goals, drives them to learn a bit about the computer and software but at the same time makes them reluctant to take time away from their own goals to learn the necessary steps or to learn a more efficient method if one exists. Consequently, Santhanam (1993) discovered that users, 'even voiced their frustration over the long time taken to accomplish this task, yet they showed no interest in learning more.'

Moreover, Santhanam (1993) revealed that an 'assimulation paradox' seemed to exist as well, where in an individual's natural attempt to understand what is being learnt, reference is made to what the individual already knows. However, any current knowledge and experiences will hinder the learning process when the analogies are not wholly appropriate for the new situation. Consequently if the current, or common, manual tasks are to be mimicked by the system, that model has to be realistic otherwise confusion and delays will result.

User satisfaction regarding the speed of interaction seems also to be linked to the actual task being undertaken. For example, users are more willing to wait longer for an answer to a 'large' transaction than during normal data entry (Miller, 1968). For although it has been argued that it is impossible for a person to accurately estimate time (Michon, 1972), it has been shown that variability of response rate influences a user's perception of the system (Miller and Thomas, 1976). Consequently, similar tasks need to take the same amount of time and status messages are required to be displayed during long periods of user inactivity.

User frustration is caused by long delays in system response and the requirement of specialist training to achieve routine tasks. It can be avoided by good system design and user participation in the design process. When operating the system, the users should be able to quickly achieve a goal which they perceive as non-trivial and genuinely helpful. This will provide the motivation and reinforcement to learn more, without which the users can quickly become disillusioned and alienated from further use of computer aids (Nickerson, 1981).

With this project, the tools which build the IDDA end-system are operated at the users' leisure. The domain expert decides on the empirical investigation to be undertaken and defines appropriate questionnaires for it. The IDDA system is then constructed from these specifications. Therefore the system can be tailored, by the domain expert, to closely match current manual procedures and the working practices of the unit. Consequently, it should be relatively intuitive to use and therefore quick to learn, with few irrelevant questions being asked.

Data can be entered directly into the computer as it is gathered or else it can be recorded and entered later. The users can adopt whichever approach they perceive as being least disruptive to themselves, their 'clients' and their colleagues.

Data analysis, however, is highly likely to occur when domain experts are not involved in any consultations and have a period of time when they are likely to be undisturbed. Therefore both the acts of building, and operating, the IDDA end-system can be determined by the users and thus can be at their convenience.

Data entry is straightforward and relatively quick. If there are any slight delays during data analysis, as the information is being collected and sorted, the user is kept informed of the current status of the IDDA system. On-line help is available, so too are various 'quick' facilities. These are all described in more detail in Chapter 4. Therefore, the IDDA system is quick, easy to learn and use and will save an investigator substantial amounts of time during data reviews and analysis.

A further crucial factor which determines whether a system will be accepted and used is the effect such a system will have on the user's work and the environment in which it is to operate. This is discussed briefly in the following section.

3.6 Effect on the user's work and environment

Gremy (1989) states that information systems should be limited to the domains where human mental abilities can be reasonably formalised (see Chapter 1). The reasons being that within these areas machines are faster, more efficient and more thorough than people. He points out, however, that any formalised model is an 'incomplete description of reality', since it lacks intuition, motivation, judgment, ethics and wisdom.

These characteristics are required, and are indeed essential, for such tasks as decision-making. It has been the inability of expert systems to encompass these features that has drastically restricted their applicability for domains requiring such tasks. If, however, the emphasis is redirected to one of assistance, as a tool, rather than one of substitution then the human users can contribute their own intelligence and a more 'complete' model will result.

Therefore there is the possibility that the introduction of computer systems could influence and change substantially the environment in which users are working, as well as their roles. This has been cited as one of the main reasons for implemented systems not being adopted by their intended users (Lings et al, 1991; Jojo, 1994). Consequently, if the impact of the system on the environment and user tasks has not been investigated, 'the net result may well be a system which either fails to properly support, or even inhibits, individual work practices and redefines individual responsibilities in a way which was not anticipated and is perhaps unreasonable' (Lings et al, 1991). Hence, resistance to the system by the intended users results. It has then, generally, taken significant changes to the original system, involving costly time delays, before users can be persuaded to reluctantly try to use it.

If, however, a system is portrayed as a tool, enabling its users to think better and use their intelligence to its highest potential, it is more likely to be accepted by its prospective users. Any system which removes or reduces the number of tiresome tasks from a user's job, and/or increases the 'interesting' aspects, will generally be welcomed.

These phenomena have been demonstrated during the introduction of many pervious computerised systems. As Rector (1989) reported, 'word processors were taken up first by academic consultants and by staff in departments which were short of secretaries. Most secretaries accepted them only reluctantly, at least at first'. In a similar fashion, 'spreadsheets were taken up instantly by departments such as catering, estates management, and the various laboratories, but that it took much longer before the finance department made extensive use of them' (Rector, 1989). Experiences like these demonstrate the difficulty of introducing a system which impinges on users' primary skills and expertise and how systems, which assist with peripheral tasks, are much more readily acceptable to end-users. Guggenheim and Whitfield (1991) agrees, 'the capacity of the computer to handle mid-stage operations removes laborious work of re-drawing, re-drafting, readjusting, re-specifying, etc. Effectively it removes unstimulating work at this stage, work of low interest and therefore low pleasure'.

Therefore the reluctance of certain users could be overcome if the emphasis is placed on developing systems which assist them in tiresome tasks and do not attempt to be a substitute for a human expert or to dictate how users should work or think. The goal of these systems should be 'to become part of the regular tools of the trade, just as the other tools which are "readily to hand" in the user's environment' (Rector, 1989).

In certain countries, user groups have been granted powers to prevent the installation of those systems which are perceived to undermine the skills and self-esteem of the workers. For instance, in Sweden a worker's council may even refuse the introduction of a tool if the council believes it will lower the quality of work-life. Furthermore, a worker's council can veto the installation of a product which it believes is too hard to learn or use (Thomas, 1983). This is because psychological stress has been linked to heart disease, alcoholism and even cancer. Therefore a system which is very frustrating to use could be considered by some doctors to be unhealthy.

France too has company doctors who have the ability to stop the introduction of a system which they believe would be detrimental to health. Germany has also passed laws requiring products to conform to strict ergonomic regulations in order for an employer to be eligible for worker's compensation insurance (Thomas, 1983). This demonstrates the importance that many Western European countries place on human factors in the design of computer systems and the emphasis they assign to the 'quality of worklife' as well as productivity.

Within medical fields, the main concern is the possible depersonalistion of medical practices. Moidu and Wigertz (1989) discovered in their studies of doctors from developing countries that these doctors were greatly concerned that the use of computers would dehumanise medicine. This fear of the effect of computers on the patient / doctor relationship has been discussed widely within the medical domain, (mainly by doctors, themselves). One argument is that 'the patient might be upset by the presence of the terminal' (Bridge and Williamson, 1981).

However, a survey conducted by Pringle et al (1984), revealed that only 17%, from 350, patients were opposed to their doctor using a computer. The major apprehensions were the aspects of confidentiality, security and the fear of 'big brother'. These concerns are not restricted solely to the medical sector, but exist over the whole area of computing and are continually being reviewed.

In addition, it has become evident that patient care is changing. Whereas physicians are worried about depersonalisation, patients themselves are demanding more information and knowledge regarding their illnesses and treatments. This is placing even more severe time pressures on doctors. Nevertheless, user acceptance and willingness to use a computer are essential for the successful introduction of a computer system. Consequently, the advantages and benefits that computers could bring, must be stressed. A 'motivating factor for the physician may well be a more satisfied patient, who has learned more about his health and disease at a lower cost, in terms of time, to the physician' (Fisher, 1985).

Pringle's (1984) results indicated that patients trusted their doctor and believed that their doctor will not allow depersonalisation to occur. Moreover, Moidu and Wigertz (1989) in their study, discovered that the doctors who feared dehumanisation had neither experience with computers nor any basic knowledge of them. Thus, Moidu and Wigertz believed that knowledge of computers and computing would place these fears in perspective. This knowledge could be achieved via education and training, but a better method would be through 'integrated system development design', i.e. having the end-users involved in every step of the design cycle. The developed system should therefore have addressed all of the problems of the working environment envisaged by the end-users (the doctors).

This current project aims to overcome many of these fears by:

- the users, themselves, designing and building the IDDA end-system. This will enable the users to adapt it to their working practices and environment,
- the users deciding when and where such a system is used. Thus, in medical fields, the actual approach adopted could in fact be different for each individual doctor. Consequently, the final effect on the patient-doctor relationship could be determined by the doctors themselves and could vary from patient to patient,
- the emphasis that the IDDA system is a tool to assist the experts by storing the collected details, enabling the records to be reviewed and analysed. This will help the experts by providing a variety of techniques with which to investigate their specialist fields.

Consequently, the users will understand:

- the capabilities of the system,
- the benefits it can bring, including the removal of laborious, unstimulating work,
- the effect of its introduction on their work and environment (which they, in fact, determine).

Thus, the users are in control of the whole situation regarding the construction, use and impact of the IDDA system rather than finding themselves in the position of being dictated to by a computer. Hence, the overall user acceptance and willingness to use the IDDA system should be enhanced.

3.7 User interface and interaction

One of the clearest lessons learned from the early expert systems is that excellent decision-making performance will not in itself guarantee user acceptance. Users must be able to communicate effectively with systems which they are supposed to use. If the interface is not liked or is perceived to be difficult to operate, then, no matter how clever the system is, it is highly likely to be ignored by its intended users. Carroll and Campbell (1989) discovered that frustration quickly results when users attempt to use computers and find that tasks which they could accomplish previously must now be re-learnt - 'competent secretaries, accountants and lawyers are - at least temporarily - returned to varying levels of incompetence until they can master "the system" '.

Furthermore, interface development consumes a significant amount of:

- the overall budget for projects, approximately half,
- the actual software that is written, at least 30-35%,
- and thus, the final development effort (Diaper, 1989; Berry and Hart, 1991).

Consequently, the user interface is a crucial component of any computer system. As Berry and Broadbent (1987) state, 'failure to recognise the mmi [man-machine interaction] needs of expert system users is probably the biggest reason for the disparity between the numerous expert systems which have been successfully developed in the laboratory and the small number which have actually made it into everyday field use.'

As the interface is so important, it should be prototyped before any other part of a proposed system and should be used as a tool for eliciting user requirements. However, the interface rarely receives such attention. For although the importance of good user interfacing is well recognised and acknowledged by both researchers and developers, 'one suspects that this is lip service rather than a seriously undertaken aspect of a project when little is done in the early stages to discover user requirements and preferences' (Diaper, 1989). Yet, to design an effective interface requires an understanding of what information is needed and how and when it should be displayed.

In addition, as stressed in Section 3.6, the working environment must be investigated to determine its impact upon the interaction process. As Guignat (1993) discovered, 'the turnover of the nursing staff may be high. Because the workload is heavy, there is no time to read extensive operating manuals, instruction cards, or help texts. Because of economic pressures on the health care system and clinical personnel shortages, especially in nursing, less time is available for in-service training.'

Consequently, the aim should be to design an interface which enables a typical user to sit down at the computer and perform a task, where it is obvious at every stage what to do next (within the context of the task). If this is achieved, the interface will recede from the user's consciousness and will be replaced by a deeper concentration on the task itself (Parsaye et al, 1989; Berry, 1994; Howarth, 1987).

This suggests that well-designed interfaces should be:

- intuitive and easy to understand,
- easy to use, with a low probability of error,
- easily mastered, requiring minimum training of the people who need to use it.

However, one major difficulty with human-computer interaction design is that it constantly changes, i.e. it is strongly influenced by the new and evolving innovations in computer hardware, software and user preferences.

3.7.1 User interface hardware

Though computers are increasing in power all the time, the actual speed and efficiency that can be achieved now relies primarily upon the human component in the human-computer interaction. 'User comfort' and ease of use are not limited purely to the content of what is being displayed but also how it is displayed. Considerable work has been done in the field of ergonomics to investigate many of the different features of display design. For example, two areas covered were:

a) the display's physical parameters, e.g. brightness, contrast, colour, flicker, etc,
b) the perceptual parameters of the material presented, e.g. character shape, size, underlining, spacing, etc.

These investigations produced various guidelines and design criteria (Oborne, 1985; Coates and Vlaeminke, 1987). The issues are important in determining how a display is perceived by an end-user and the effects it has on an end-user's performance. Appendix B describes some of these aspects in more detail.

The physical display is not the only aspect of human-computer interaction that requires investigation. There are other factors, such as determining the best devices for the operator to enter information. The most common method is the QWERTY keyboard though others do exist, such as touch displays, light pens, joysticks and the mouse.

The normal QWERTY keyboard has been in use since the beginning of this century. It was designed originally to conform to the mechanical constraints of early typewriters: the apparent haphazard arrangement of letters was developed to slow typists down to prevent the jamming of the keys. This is an example of how established practise can become accepted within new technology for entirely the wrong reasons (Oborne, 1987). Noyes (1983) however did discover one saving factor: the QWERTY keyboard design does evenly distribute the workload assigned to each hand and may therefore reduce user fatigue.

Other kinds of keyboard have been proposed over the years, for example the Dvorak keyboard (patented by Dvorak in 1932) which has all the vowels and the most used consonants placed on the second row, and the alphabetic keyboard where the keys are arranged from A to Z. The pros and cons of these two different keyboard layouts and of a selection of the other input devices are discussed in Appendix B.

Unfortunately, there have been few comparative studies investigating the best type of device for particular circumstances and user populations. Nevertheless, it is essential that the input device is compatible with the type of dialogue selected for the computer system. A pointing device, such as a light pen or a mouse, is of little use when a natural language interface is to be used. However, when hybrid interfaces are utilised in one system, for example menus and form filling screens, the difficult question arises as to whether multiple input devices should be integrated or would they merely confuse the user?

In a study undertaken by Berry and Broadbent (1987), the experts preferred to use just the keyboard rather than to switch to the mouse for selection and back again, even though the mouse has been accredited with being the most efficient of the pointing devices. Maclean et al (1985) reported similar results and concluded that there seemed to exist a 'task boundary' which had to be overcome during the transition from the keyboard to the mouse or vice versa. This required an amount of cognitive resources and therefore was more difficult to accomplish than if no switching was necessary.

In addition, it seems that if a choice of input device is given to users during a task, users will take longer to accomplish the task. This was discovered by Olson and Nilsen (1988) in their studies of users operating two packages; Lotus 1-2-3 and Multiplan. They showed that, in addition to the time taken for motor movement and the cognitive processes required to move between input devices, time was also required by users to select between devices offered. Lotus gives users two methods by which to select a cell on the spreadsheet, either by typing the cell co-ordinates or by using the cursor keys to point to the appropriate cell. Multiplan on the other hand has only one method, using the cursor keys. Olson and Nilson (1988) discovered that although the time to enter the formula was the same, users of the Lotus package took, on average, an additional 1760 milliseconds in deciding which method to use before starting an entry, i.e. the time to start a formula in Lotus was 4.63 seconds whereas in Multiplan it took only 2.87 seconds. This therefore suggests that a choice between methods in human-computer interaction is a complex cognitive task requiring several cognitive steps to be executed. Consequently, an important factor during interface design is to ensure that the cognitive load being placed on a user, by both the interface and the input devices, is not too great otherwise it will effect the real tasks a user is attempting to complete, i.e. decision making or data entry.

There is evidence that by reducing the number of actions required to complete a task, i.e. simplifying the structure of the task (Norman, 1988), the amount of working memory required is reduced, which in turn reduces the number of errors. Even skilled operators make errors and it has been recognised that some of these errors arise from an interface design that requires too much to be held in working memory (Olson and Olson, 1990). Therefore, 'an estimate of cognitive complexity for information processing has to be added with the objective of enabling the designer to troubleshoot the design for human limitations of working memory' (Sutcliffe, 1990).

Cognitive modelling has recently become one of the central issues in the design of a system. These models simulate the cognitive processes and task knowledge of users to enable a designer to determine various aspects of an interface's usability, for example, identifying when user errors are likely to occur (Gugerty, 1993; Olson and Olson, 1990). Consequently, early interface design can be quickly evaluated and improved with these models.

The belief is that a system will be easier to learn if the amount of knowledge required to carry out a task is reduced. To do this, there must first be some indications as to the interaction requirements needed to accomplish the task. Hence, the ability to predict how users interact with the proposed designs is a useful tool for the system designer. Card et al (1983) in their GOMS (Goals, Operators, Methods, Selection) model attempt to do this by characterising the knowledge necessary to make effective, routine use of software tools, such as a text editor.

The GOMS model is used to predict the time it takes a skilled user to execute a task. This is based on the composite actions of retrieving plans from long-term memory, choosing among alternatives, keeping track of what has been done and what needs to be done, and executing the motor movements necessary. The model assumes that routine cognitive skills can be described as a serial sequence of cognitive operations and motor activities. The model could, therefore, be used to make performance predictions for expert users carrying out routine tasks.

Olson and Olson (1990) reviewed the estimates of operator times from a number of studies, which had used the GOMS model, and found that there was a fair amount of agreement between them. For example, a keystroke took about 230 ms and retrieving an item from memory about 1250 ms. They then used these estimates, and GOMS modelling, to accurately predict user performance in another task. Consequently, methods such as GOMS can be of value to system designers, especially for quick evaluations of early interface designs.

It must be noted, however, that such approaches are not fully complete. For example: they model only sequential tasks, they require interruption free environments, they do not capture the impact of fatigue or stress on the times and errors associated with performance, nor do they include an assessment of the user's perception of the acceptability of an interface. 'Perceptions of whether the system's functionality is actually what the user needs and the ease with which the system can be learned, all contribute to the acceptance of a piece of software and its eventual regular use' (Olson and Olson, 1990).

Cost of hardware, of course, is another factor in interface design decisions, especially in medical fields. Many medical applications begin as research efforts and, therefore, tend to be strongly independent in their choice of computer systems. This results in a disorganised array of hardware and software, which often inhibits the transportation of systems from one centre to another. Therefore there is a reluctance for institutions to share in the development costs of a system.

Friedman and Gustafson (1977) found 32 articles presenting applications of computers to medical problems. For 51% of the projects reviewed, the work had either been abandoned or temporarily suspended. 19% were in routine use at their medical centres. Friedman discovered that in almost every case where the project had been abandoned, it was because the project had never become cost effective and, when the research funding ran out, the hospital could not resume the funding. Therefore even though the cost of hardware is dropping, the cost of the whole system is still a crucial factor in determining whether medical systems will be completed. Cost should not be assessed purely in the monetary terms of purchasing the hardware and software. It should include the total financial expense, i.e. the time and the effort required to design, develop, test and maintain the system as well as the purchasing costs.

One of the objectives of this research is to develop a general methodology to permit as many fields and experts as possible to benefit from computerising their investigative procedures. This has meant that the cost of the solution had to be kept to a minimum, thus the choice of possible hardware products was effected. Moreover, it was decided that if the hardware selected could also fulfil other roles in the domain, this would be highly desirable and increase the likelihood that the expense could be justified. An IBM compatible PC appeared to be the best option as such a system is readily available, hence cheap, easily maintained and can be upgraded, if required. Furthermore, there an abundance of software already available for such machines.

Expense also ruled out specialist input devices, such as touch screens. As the literature indicated, multiple devices could confuse the user, slow the interaction speed and cause errors, especially if the operator is a naive computer user. Consequently, the decision was to select one method. It should be able to cope with all the input techniques used in the various interfaces. Therefore the keyboard seemed to be the obvious choice. Not only is it integral part of a computer, thus included in the original price, but it allows entries to be made in long-hand as well as through one keystroke selections. In addition, if a user has had any previous experience of computing, or typing, they would have used a keyboard. This skill could therefore be transferred.

Consequently, techniques such as GOMS were not used for the decisions regarding hardware. Monetary costs were the influential factor. Moreover, GOMS was not appropriate, in this project, for analysing the initial interface designs. The main reasons were:

- it modelled expert users rather than naive users,
- it required sequential tasks rather than parallel processing, which could occur, in certain scenarios with the IDDA system, if the data gathering and data entry tasks were being undertaken concurrently by the user,
- it required an interruption free environment,
- it required users who made few errors.

Therefore for this project, the interfaces were developed through the more traditional prototyping approach. Although this method has been deemed to be expensive and time-consuming, with this project a test-bed already existed in the operational Knee Ligament system at the LRI. Therefore that system was reviewed initially and the IDDA system interfaces then evolved through an iterative cycle.

The only unusual aspect of the hardware configuration was the decision to use two monitors. This was to enable the main terminal to be used solely as the entry screen and not to be cluttered with help text or verbose descriptions. Any help information, examples or extra details could then be shown on the second monitor. Such a set-up enables the user to have, at hand, a reference screen for assistance without covering the questions attempting to be answered on the main monitor. Thus users can compare the correct format given in an example whilst they are typing in the information. Help is available during both the IDDA end-system construction phases and the final IDDA system operation. The reasons for this decision are discussed in more detail in Chapter 4.

Hardware requirements should only be part of a full analysis into the design of a system interface. Other factors such as user characteristics, consistency, display and dialogue methods all influence the impact and user acceptance of an end-system.

3.7.2 User interface software

Interface requirements differ depending upon the users' previous computer experience, their abilities, their background and the tasks they need to accomplish. No single design can satisfy all users and all situations. At the beginning of the design process, therefore, an analysis of the proposed users, their characteristics and the task domain must be undertaken.

There is a belief that users construct mental models of a system to assist them in performing tasks and correcting errors when they use an application (Trumbly et al, 1993; Berry, 1994). These models are formed by using analogies or metaphors of past experiences. Thus, a good design strategy is to use as a metaphor something that the user already knows. 'The closer the system model is matched to user expectations, the more easily and quickly user learning takes place' (Gerlach and Kuo, 1991). Hence, mental models can be used to assist in user learning, performance and system design.

Staggers and Norcio (1993) discovered that if help is given to assist users in building conceptual models during training, they seem to be capable of understanding the system better and outperforming others in complex tasks. The conceptual models seemed to act as knowledge frameworks in which new knowledge was organised. However, Douglas (1982) and Lewis and Mack (1982) warn that the analogies must be accurate. They reported that their subjects tended to generate a typewriter model when they learned to use a text editor. This then led their users to possess and exhibit the types of errors and misconceptions about the behaviour of the text editors, which were consistent with this analogy.

Therefore, to ensure that a system is used effectively, users must have an adequate conceptual model of what the system does and be able to interact with it. Cognitive compatibility can briefly be defined as the degree to which the model of the task presented by the interfaces conforms to the corresponding expectations of the user (Parsaye et al, 1989; Berry and Hart, 1991). It is highly desirable for a system to be perceived as being 'natural' to use so that a user can concentrate on the problem being solved rather than on how to operate the system. However, realism 'cannot be easily achieved because technological restrictions limit the choice of dialog style and impose rigid syntax rules and recovery procedures', (Gerlach and Kuo, 1991).

Guignat (1993) used mental models during the design and construction of his system, 'in most cases, nurses and physicians have no computer experience. .... Working towards a simple model in the user's mind was considered more important than reducing the number of keystrokes required to access a given function to an absolute minimum. Having formed a model of how the system operates, the user can extrapolate how a particular function might work. If the system is consistent, the user's prediction will work, the system will be perceived as easy to use, and user acceptance and satisfaction will increase'.

However, users vary in their past experiences and skills, especially with regards to computing. Consequently a number of categories have been established as a mechanism for grouping users together to permit the specification of general requirements.

Shneiderman (1987) describes the requirements for three of these groups:

Novice or Naive users:

> This group is assumed to have little knowledge of computer issues. They may be anxious about using computers and this will inhibit their ability to learn. To overcome these limitations: restrict the vocabulary used to a small number of familiar, consistently used terms thus developing a user's knowledge of the system. Also keep the number of input possibilities to a minimum. Novice users should be quickly able to carry out a few simple tasks, thus building confidence, reducing anxiety and enabling them to gain positive reinforcement from success. Specific error messages should be provided when errors occur.

Intermittent users:

> Many people may be knowledgeable but intermittent users of a variety of systems. The burden on memory will be lifted by simple and consistent structure in the command language, menus, terminology and sequences of actions, meaningful messages and frequent prompts. Protection from danger is necessary to support relaxed exploration of features or attempts to invoke partially forgotten commands. These users will benefit from online help screens to fill in any missing pieces of knowledge.

Frequent users:

> The knowledgeable users are thoroughly familiar with the various aspects of the system and seek to get their work done rapidly. They demand rapid response times, brief and less distracting feedback, and the capacity to carry out actions with a few keystrokes or selections.

These characteristics can be refined for the task environment and then used as the basis for selecting and designing the correct type of interface for the proposed users. If only one class of user is expected, the design process is much easier. It becomes a far more complex task when multiple classes of user are expected.

The problems arise because 'software written for novices may be too cumbersome for its now expert user. Similarly, software written for the expert may be nearly useless until expertise is painfully acquired' (Vaubel and Gettys, 1990). Carroll and Rosson (1985) agree 'it would mean little if [only] a sophisticated programmer were able to understand the information presented on a given display panel of a word processing system intended for a secretarial population.'

Shaw and Woodward (1988) raised a question regarding expert system development. Since expert systems are developed in close consultation with experts will a novice be able to accurately understand the questions posed by the system. Vaubel and Getty (1990) continue this theme by pointing out that help text appropriate for experts is generally incomprehensible to novices. Comprehensibility is essential for successful interaction since without knowing what is being requested, a user will not know how to respond.

Montazemi (1991) discovered that novices preferred presentations which provided visual aids to help them comprehend the decision problem and to extract the information required in their decision making. This supports the belief held by Vaubel and Gettys (1990), that novices spend most of their time searching for information and frequently need help. Experts, however, rarely need help except to refresh their memory about the details of infrequently used commands.

With these major differences in user characteristics, and therefore user requirements, there has been a move towards adaptive interfaces for those environments which are trying to accommodate a number of user classes. These interfaces enable the displays and the help presented to a user to be altered as a user's experience grows. This area of interface design is not covered here but details can be found in the following papers: Innocent (1982), Greenberg and Witten (1983), Fowler et al (1987), Williges et al (1987), and Vaubel and Gettys (1990).

Since this research is mainly concerned with naive computer users and their requirements, the rest of this section will concentrate on the dialogue and interface design needs relevant to this particular group.

The dialogue facilities of a system must match the communication needs of users and the constraints of the task environment. As Trumbly et al (1993) reported, users can perform better, in terms of error ratios and task performance, when the software interface matches the skill and knowledge levels of a user. Correct display and dialogue designs are therefore essential to help system learning and acceptance. The difficulty with medical fields is the absence of a standard medical language. It is important to permit the doctors to use their own vocabulary and not to force them to adopt a different style. A system which looks and feels familiar is more likely to be accepted by its prospective users (Sutcliffe and McDermott, 1991; Berry and Broadbent, 1987; Hayes-Roth and Jacobstein, 1994).

The most commonly reviewed areas of dialogue design are: consistency, naturalness, non-redundancy and supportiveness. Each is important, however since they have already been discussed in detail by other researchers they are explained in Appendix B.

Various types of dialogue exist. The four basic types are: question and answer, menus, forms, and, command line. They are not really distinct but are all variations of the question and answer structure, though each has its own strengths:

a) a **menu** is displayed to a user before a selection is made and thus provides a user with first level help at the same time. This structure would suit an inexperienced user or where a limited range of values existed.

b) **form filling** is a sequence of questions displayed at the same time but answered one by one. An advantage of form-filling is that it can support the users during the process of assembling their inputs to the computer (Frohlich et al, 1985). This method is most applicable when a standard sequence of collecting and entering information is to be followed.

c) the **command line** is similar to question and answer but it also enables a user to answer ahead, i.e. it allows a number of answers to a series of questions to be entered as one entry. The input values and the sequence of entering the data have to be remembered (Kantorowitz and Sudarsky, 1989). This means that it is not ideal for every situation and should not be viewed as the ultimate goal by system developers. It can, however, be appropriate for experienced, frequent users where a limited amount of data must be entered before an action occurs, i.e. a fairly flat hierarchy of task processes.

d) with the **question and answer structure** itself, questions are displayed and the answers entered one by one. This can be used at any time, but is most appropriate if there are too many options for a menu, or the sequence of entry is too complex for a command structure, or if the next question to be asked is dependent upon the response to the current question.

An understanding of these characteristics is useful in the design of systems, particularly when determining which structure will be most appropriate for which environment. In fact, Gerlach and Kuo (1991) claim that, 'the system model, when designed in accord with user perception of how tasks are conducted, may suggest the dialog style. For example, the 'form' style is the natural choice for a system involving database inquiries because forms are widely used for storing data manually and, as a consequnce, become the metaphor for that system'. It should be noted, however, that in general different parts of a system have different requirements and, therefore, require different dialogue structures. Rarely, does a whole system contain just one type of dialogue. In most systems a hybrid approach is adopted, i.e. different dialogues are used in the various sections of the system where they are most appropriate, thereby combining the strengths that each approach offers.

An inexperienced, or infrequent, user requires more assistance and explanation in instructions and messages than an operator who frequently uses the system. With a command language interface the user must know and understand thoroughly the language. Thus, as there is less structure and assistance, user errors can easily occur. Even if the command language used is a restricted natural dialogue, questions have been raised over a user's ability to adapt to a subset of language (Garg-Janardan and Salvendy, 1988a).

With the other dialogue styles, operators need not become familiar with a command language itself. Consequently, they can use the new system after a very short period of time. With a menu mode, a new user can explore the operations provided by the system simply by browsing through the menus. Menus are also popular because they have the ability to force answers into a limited number of choices, thereby reducing the possibility of input errors and long-term memory requirments. The system guides a user, thus enabling a user to concentrate on the task in hand (Simpson ,1982; Trumbly et al, 1993). Moreover, Kantorowitz and Sudarsky (1989) discovered that if help is provided for every menu option, a user can operate the system without ever needing a manual. Consequently, such characteristics have been the main reasons why novice and intermittent users often express a preference for a menu style dialogue (Heydemann et al, 1991; Santhanam, 1993).

Van Hoe et al (1990) carried out a study to determine the best method for a menu system, i.e. either the breadth or depth structure. 'Breadth' represents menus with a lot of options in each menu and with few sub-menus or layers. 'Depth' is the opposite, i.e. it refers to menus containing few selections but with numerous layers. Van Hoe et al concluded that good user performance is enhanced when hierarchical menus are organised by 'breadth' rather than by 'depth'. In addition, they indicated that an escape feature facility should be included.

The ability to escape out of an incorrect choice is essential. Carroll and Carrithers (1984) found from their studies of novices learning computer systems that a great deal of time was spent recovering from inappropriate menu choices. However, presenting only the possibilities that are appropriate to a user's current state would be one method which could help prevent this type of error.

A number of studies revealed that medical end-users found the necessity of typing all the relevant information in long-hand to be too cumbersome (Hudson and Cohen, 1985; Safran et al, 1991; Shiffman et al, 1991). Many had become accustomed to the use of home computers where application programs were menu oriented. They therefore preferred this type of interface where they could select an entry, even though it can be a slower method when compared with the others, e.g. the command language interface. The results of MacLean et al (1985) also support this view that there is a strong and consistent bias in favour of the menu method for entering data.

Shneiderman (1987) gives a comprehensive review of the various dialogue styles. One of the main aims in designing and developing an interface is to ensure that users can interact with the computer in such a way as to achieve a task without wasting time worrying about how the actual communication will take place. As Meadow (1970) states, 'time spent thinking about how to express an idea to a computer may be time away from solving a problem. What is worse, it may actively interfere with problem-solving.'

The interface design has been shown to have a significant influence on factors such as learning time, performance speed, error rates and user satisfaction. Though a good interface can lead to significant improvements, poor designs can actually hold users back. Without a clear understanding of user needs and requirements, system development may fail to produce crucial capabilities, resulting in a system being built that has limited utility.

The above concerns were reviewed when designing the user interface and interaction mechanisms to be used for this research project.

The operators are highly likely to be naive or novice computer users. In addition, they are likely to utilise the tools and the statistical techniques of the IDDA end-system only intermittently, whilst using the data entry and review sections of the end-system frequently. Furthermore, time will always be at a premium and interuptions common-place.

By enabling the experts to specify the data entry questions on which the IDDA end-system is built, the belief is that the comprehensibility of the system should be enhanced, especially if the experts also have the ability to define their own help which can be attached to any question. Furthermore, as both the tools and the analysis section of the IDDA system, will only be used infrequently, the decision was to ensure that there would be extensive in-built help with examples available to the user on request.

With regards to the dialogue styles, it was evident that no one style would suffice. From the literature, the command line dialogue seemed to be unsuitable for the intended users of the tools and the IDDA system. At the present time, the graphical user interface (GUI) style also appeared to be inappropriate for both, due to the type of users, tasks and environments that are likely to be involved in developing and utilising IDDA systems.

With GUIs, users select objects and actions and decide on the order of tasks. Consequently, there is no implied sequence of activities. However, it has been shown that naive or novice users prefer simple, restrictive interfaces where tasks are undertaken by following fixed sequences of operations (Wright, 1994; Trumbly, 1994). They need considerable support and prefer interfaces that require a minimum amount of cognitive processing to operate (Nilsen, 1992). Moreover, as already explained earlier in the chapter, these particular user groups find multiple windows confusing and find the necessity of having to operate a mouse frustrating. In fact, these could be some of the reasons why menus continue to be utilised in the lion's share of application software despite the availability of GUIs (Trumbly, 1994).

The tasks which an IDDA system undertakes are by their nature very structured and, thus, they can be modelled well by using a combination of the menu, form-filling and question and answer styles. These styles are also good at providing the additional support required for users working within highly stressful environments, where users are frequently interrupted and have very limited amounts of time to complete tasks.

The aspects of consistency, naturalness, non-redundancy, supportiveness were also major considerations during the design. In addition, speed of operation, learnability, reduction of errors and prevention of user-overload were also influential in the decisions taken over the interface design and software operation. If the users could not use the tools, they could not build the end-system. Consequently the tools should be easy to learn, quick to use, robust in the case of errors, and cheap in terms of hardware and software. Ultimately, the envisaged users, their tasks and their environment were the driving forces behind the selection of hardware and software and the design and prototyping of both the tools and the IDDA end-system. The next chapter will describe in more detail the actual design and development decisions taken.

Chapter 4

The Design of the Tools and the IDDA End-System

4.1 Introduction

This research is to investigate whether a general methodology can be devised to enable naive computer users to develop IDDA systems for their specialist domains. The IDDA system assists users in investigating their fields. As Chapter 2, Section 2.5.1, explained, these investigations tend to follow an empirical approach which involves:

- conducting an experiment,
- analysing the collected data and comparing the results,
- inducing (or refining) a hypothesis from the analyses,
- and, using this hypothesis to predict the results of future trials, thus iterating the cycle.

The experimental stage, itself, involves recording: the initial state of the objects under investigation, the changes made to the conditions or objects, and the final state of the objects after a pre-determined time period. This experiment stage is concerned with data collection and data storage, thus it can be mapped in an IDDA system by the data entry section and the database itself. The analysis stage of an empirical investigation can be incorporated in an IDDA system through the data review and the statistical analysis facilities. The last two stages of the cycle are undertaken by domain experts as it is in these tasks that the strengths of human reasoning and problem-solving lie.

This chapter discusses the design decisions involved in constructing the experimental stage for such an IDDA system, and their subsequent effect on the development of the computer-based tools. Chapter 5 reviews the decisions involved in the analysis stage. However, before any development began, the problems experienced by previous system developers had to be identified (see Chapter 3). As the literature showed, a system has to operate around users and not users around the system. The environment and the pressures placed upon users must be considered otherwise the acceptability and willingness to use the system will be greatly diminished. These issues have been quoted as the reasons why so many of those early systems failed to be implemented and used outside research or development environments. The critical factors can be summarised as:

a) the wrong goals were pursued by the programmers,
b) the end-systems did not follow the normal working practices of users,
c) the lack of consultation with users during the design and development stages,
d) the apprehension of the users and the perceived threat of computers,
e) the amount of time and effort required to develop the end-system and then to run it,
f) the adverse effect on the users' work and environment,
g) the poor user interface and the lack of support given to a user by the system.

The following sections outline the approaches used in an attempt to minimise these problems during this research study. To assist in the understanding of both the medical procedures and practices which are commonly undertaken during a medical investigation, and the typical characteristics of a medical computer user, the orthopaedic group at the LRI were consulted and relevant medical literature was reviewed. Though specialist medical fields have been used as information sources and test-beds, it is intended that the generic IDDA system design will be appropriate for any specialist domain which undertakes similar empirical investigations.

4.2 Medical procedures and practices

Typically, within many specialist fields, there exists at least two and normally three assessment stages during an investigation:

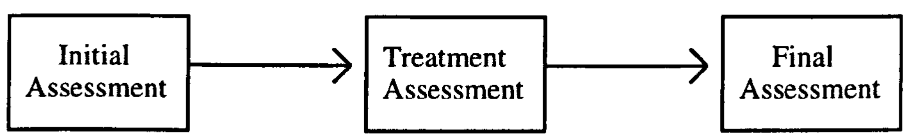


Figure 4.1: A three-stage assessment process

In certain ailments, for example cystic fibrosis or asthma, there may only be two stages:

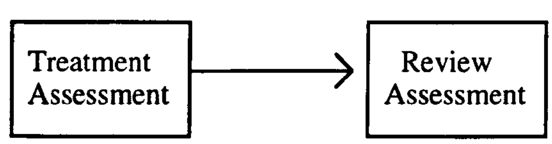


Figure 4.2: A two-stage assessment process

Loops back to previous assessments following a relapse, failed surgery, or reinjury, will be dependent upon the ailment concerned. For example, in a 3-stage assessment process for knee ligament injuries, the loops back for re-assessment purposes might be:

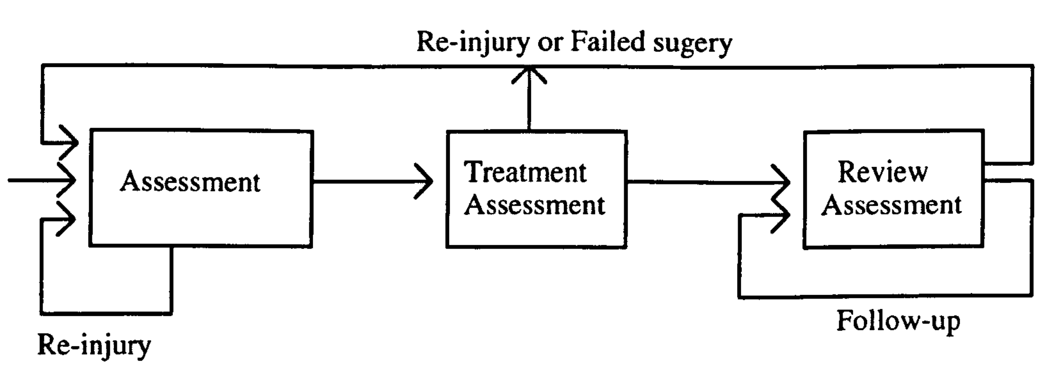


Figure 4.3: A 3-stage assessment process for knee ligament injuries

The exit points from the assessment process will be dependent upon the investigation being undertaken, e.g. the number of follow-up visits, patients requesting removal from the study, etc. Any computer-based system must be capable of dealing with all possibilities and with linking together all the information of a particular patient. An assumption made at this point is that there will not be more than one assessment at the same stage carried out on one patient on the same day. If this situation ever arose, it is assumed that only one set of data would be required to be retained.

Domain experts design investigations, thereby determining the number of stages involved. Moreover, they will be specifying the data collected during the investigations, the procedures required to obtain this data, and the sequence of these procedures and questions. Experts are able to define these details, as they are aware of the theories and hypotheses they wish to test and analyse. Therefore, to permit accurate statistical analyses to be carried out, pre-defined, set data must be collected, i.e. standardised assessments must be used during the investigation. The impact of introducing standardised assessments in areas where they currently do not exist should be minimal, especially in medical fields where patient records are commonly kept. As Rector (1989) points out, 'most consultations follow a stereotyped script, [though] few systems take full advantage of this fact'.

4.2.1 Medical records

Without rigorous recording of patient details and the use of standardised questionnaires and assessment procedures, statistical analyses cannot be conducted on the information gathered in an investigation. Hence, hypotheses cannot be tested nor theories be produced. Nor can previous cases be reviewed to assist in subsequent decision-making. As Gage (1993b) points out, 'ill-equipped to cope with quantitative problems ..... physicians labour in a bewildering environment in which the evidence that they should be using to make decisions is either missing or uninterpretable'.

Consequently, a standardised recording mechanism is crucial if valid and worthwhile investigations are to be undertaken. Moreover, these patient assessments should be more precise and detailed than merely reducing all observations into a single statement, which commonly occurs during diagnosis. This approach fails to convey the specific characteristics of individual patients. Hence, assessments should focus on observations and collect clinical signs, symptoms and test results, e.g. laboratory data and patient scores. In this manner, the information in the record is understandable to other physicians as well as the patient's current physician. The clinical data therefore becomes structured and measurable, allowing accurate and methodical studies to be undertaken and more reliable results and conclusions to be drawn.

Furthermore, studies by de Dombal (1988) (a qualified doctor) and by McDonald (1976) concluded that the introduction of standardised assessments would be highly beneficial as they force doctors to adopt the traditional approach when taking patient case-histories, i.e. talking to the patient carefully, listening to what patients have to say, defining the terms used by patients very carefully and, thus, recording a thorough case-history.

Doctors seem to prefer a structured approach since they work in an environment where external interruptions are frequent. Studies of medical auditing (Donabedian, 1980) have shown that doctors often omit simple checks and make simple errors despite adequate knowledge. This is due to interruptions, tiredness or haste. McDonald (1976) revealed that there seemed to be no association between the level of training or perceived skill of the doctors and the frequency with which they made routine mistakes. Mistakes, McDonald (1976) believed, could be easily audited and probably avoided with the help of a relatively simple computer system and structured, standardised questionnaires.

Standardised assessments could also lead to the evolution of a universal language, as well as allowing for investigations into the clinical procedures and tests commonly undertaken. This could result in a more formalised field, which would reduce confusion and misunderstanding over the terminology used, and prevent the wasting of valuable time, money and effort in carrying out tests and examinations that are inappropriate or inaccurate for a particular situation or scenario.

Adams et al (1986) discovered, in their study of the impact of three recording methods used during the determination of appendicitus, that the use of structured forms alone improved initial diagnostic accuracy from 45.7% to 56.7%. When these were computerised on a PC, there was a further rise to 64.8% and when feedback was also provided, an accuracy averaging 68% or higher was achieved. Data relating to bad surgical errors, appendix perforation rates and admission rates also showed similar beneficial indications. In addition, there was a reduction in the length of stay in hospital and the number of special investigations ordered and carried out. It was estimated that all of these would result in a revenue cost saving of £748,000 over two years in the project hospitals. Adams et al (1986) calculated that on a national basis, this equated to an average of £23 million a year saving to the NHS. With direct costs, the savings were estimated at £210,000 over two years in the project hospitals and, on a national basis, £5 million a year to the NHS. Adams et al (1986) concluded that these savings should be set against the cost of the system (which, in this case, was £2,500 for hardware and £500 per year maintenance and service).

Consequently, as Bradbury (1991) remarks, 'progress at this exciting time in the history of medicine is dependent upon the development of a standard for medical records. This does not mean a standard computer program or standard style of computer, but a standard framework within which medical professionals record information. Standardisation of the format of information, rather than hardware or software, will open the floodgates for an information era in medicine'.

4.2.2 Medical users

Burgess (1991) classified physicians as: intelligent, computer naive, busy, and cautious of computers. It was evident from reviewing the LRI scenario that these characteristics were indeed accurate, even though appearing rather limited. For example, it was also apparent that junior doctors regularly rotated positions within a department and hospital. In addition, in computerised settings, the entry of patient details was often batched until the end of the day or, when an emergency occurred, for a few days. Consequently, knowledge of these user characteristics influence the design of the application interfaces and the selection of the interaction styles that will be adopted in the computer system.

Hence, when the users are intelligent, basic assumptions can be made, e.g. that they will understand that their actions at the keyboard will produce a reaction from the computer, etc. If they are computer naive, however, the interface must be carefully designed to ensure that they know what is expected of them and what the capability of the computer is. If they are very busy, they will have little time to learn new skills, read large manuals, or complete extensive tutorials. Therefore the interface must be intuitive and 'natural' to use. If they are cautious, it is generally due to the lack of computing knowledge or experience as well as the perception that they currently carry out their jobs satisfactorily without a computer. Hence, the system must not appear to require any extra work whilst quickly producing obvious benefits. It should integrate easily into the users' daily working practices and relieve the users of tasks which are laborious, mundane and slow. If the users rotate jobs after a period of time, the user characteristics are unlikely to change as new users fill the vacancies, i.e. naive users do not gradually progress to become expert users. Finally, if the users utilise the application intermittently, tasks which appear to be routine at one point in time become non-routine at a later date. As Santhanam (1993) observed in their study, users had difficulty in remembering commands and actions when they had gaps in their use of software. Thus, the system must provide more user assistance and support than normal. These requirements are similar to those which will be demanded by applications which are utilised in an environment where there are frequent interruptions.

Many of these concerns could be alleviated and addressed if the interface appeared 'natural' to the user, i.e. if the domain terminology, question order, expected responses and layout are all tailored to the users' traditional mode of expression. Moreover, if the users can build an appropriate mental model of the computer application, the users' conceptual barriers regarding computers and the application will be reduced and user acceptance will be enhanced. As forms are the natural method used to record data, and are used by a large number of individuals in their daily lives, the adoption of a forms approach for a data processing activity would enable the users to construct a suitable conceptual model.

Outside a form-filling situation, the literature has shown that naive and occasional users prefer menu-based systems as they can select options from a list rather than enter free-format text (Trumbly et al, 1993; Shneiderman, 1991; Kantorowitz and Sudarsky, 1989). In addition, it has been stated (Safran et al, 1991; Guignat, 1993; Bradbury, 1991; Parsaye and Chignell, 1993) that they favour an interface which:

- requires limited amounts of typing,
- uses few abbreviations and codes, even those which are common within their own domains,
- provides type and range checking,
- involves minimal complex navigation,
- utilises common response values, i.e. default values.

Although the tools and the IDDA end-system have different objectives, (the aim of the tools is to develop an effective and efficient end-system whilst the IDDA system needs to collect, review and process the stored data by appropriate means) they both interact with naive computer users. Consequently, the underlying design specifications and the user acceptability issues are similar. These factors are crucial in determining whether a system will actually be used after it is implemented. Hence, they were influential in the design phases of both the methodology and, subsequently, the tools.

Some of the more important issues that have been highlighted include:

a) ease of use - screens should be consistent;
help must always be available;
the terminology used must be understandable to the user;
little training must be required;
there must be error recovery.

b) speed of use - a limit to the number of errors that can occur, by the presentation of only the options available at that time;
help on request rather than vast quantities of text on each screen.

c) users contentment with the system -
feeling of confidence with the computer;
uncluttered displays;
users never left not knowing what to do next;
selections should cover all possibilities, e.g. users should not have to adjust their input to suit the system.

If an interface can be built which can satisfy all of these issues and contain these features, busy, intermittent, naive computer users, i.e. physicians, should be able to be accommodated.

4.3 Design of the IDDA end-system

Consequently, from the research within specialist medical domains, it is evident that it is generally accepted that medical investigations should consist of 2 or 3 stages and involve the use of standardised questionnaires. In addition, human-computer interface styles appropriate for medical users have been identified and factors that have been deemed crucial for the implementation of computer systems within a medical arena have been summarised. This knowledge can therefore be combined and utilised during the design of an IDDA end-system. However, there is no way of knowing in which specialist field an IDDA end-system is to operate. Consequently, how can an interface be developed that contains the appropriate dialogue structure and terminology? The answer is to allow users to determine the interface. The end-system is to assist in empirical investigations, therefore data needs to be gathered and stored.

Typically, when domain experts initiate new investigations, they will conceive the questions that require to be answered. In a thorough study, they will formalise these questions into standard questionnaires. In this act, they are:

- specifying the number of stages involved in the investigation, i.e. 2 or 3,
- devising the questionnaires to be used at each stage to collect the required details,
- determining the order of the questions and medical test procedures,
- defining the responses expected, through the asking of the questions,
- using relevant specialist terminology for the domain and tasks,
- ensuring that the overall investigation, including the medical tests, can be undertaken within the working environment.

Hence, if these initial proposals can be computerised, the number of databases required in the IDDA system will be known, i.e. 2 or 3. Moreover, the terminology and the layout to be used for the data entry interfaces will have been specified, i.e. the questionnaires. However, the type of response expected for each question and any non-sequential form navigation would still need to be defined before a system can be built.

Nevertheless, if this approach is possible and can be adopted, those factors listed in the Introduction (Section 4.1) as being critical to the development of computer systems can be addressed. For example, domain experts determine the goals of investigations and the processes used to achieve these goals. They are aware of both the local working practices of the unit and the daily demands placed on the members of the unit. Therefore, they can automatically accommodate these issues. They are knowledgeable of the correct terminology, the most appropriate sequence of questions and tests, and they can provide additional assistance to a user by listing the available responses to specific questions.

The details from which an IDDA system can be constructed are therefore provided by the individuals who will be using the end-system. Since the IDDA system will be following similar procedures adopted by previous manual investigations and will be adopting the traditional approach of using forms for the collection of data, a conceptual model of the IDDA system can be easily constructed by all of the users. This will assist in allaying the phobia and the apprehension commonly experienced by medical computer users. With respect to the time and effort taken to develop the IDDA system, these are discussed in more detail in the next section. However, the time and effort required to run the IDDA system is chiefly dependent upon the length of both the questionnaires and the expected responses.

In addition to addressing those factors listed in the Section 4.1, this development methodology would avoid most of the difficulties associated with the current knowledge acquisition techniques, since it does not require experts to explicitly state their expertise. It merely needs them to devise the questionnaires for the proposed investigations, i.e. the data and test results they wish to collect.

Furthermore, computerising typical manual data collection techniques combined with interfaces which simulate tasks normally confronted by the physicians, creates sufficient familiarity so as to minimise the amount of training required, since the users will already have a substantial understanding of the end-system, the questions being asked and the responses to be entered. These features, along with the ability of the users to work when it suits them, should aid user acceptance and commitment to utilising the IDDA end-system.

4.4 Design of the Tools

The factors that influence the design of the IDDA system will ultimately effect the design of the tools as well. Not only in terms of the characteristics of the users who will be operating both the tools and the IDDA system, but also in terms of the decisions regarding the interface and interaction styles to be adopted by the IDDA system which will be constructed, in the main, by the tools. Consequently, these design decisions must be reflected not only in the tools' construction of an IDDA system but also in the design of the tools themselves.

These design decisions resulted in a basic layout being used for various screens involved in the tools (see figure 4.4). The background colour is dark blue except for the 'window', which is black. There is a white border around the 'window' to separate it from the rest of the screen. All function key labels are in cyan except for 'F1 - Help' which is in white. Any other information displayed in the top section or questions displayed in the lower section are in yellow whilst information appearing in the 'window' is in green. Error messages are displayed in white and appear on the bottom line of the lower section. The background colours were selected for their depth. The foreground colours needed to be attractive and to contrast with their backdrop in an appropriate manner. They were also used to signify the relative importance of their messages. A 'window' was used to focus a user's attention to the information appearing in this section of the screen. This was important since all the questions in the lower section of the screen refer to this information.

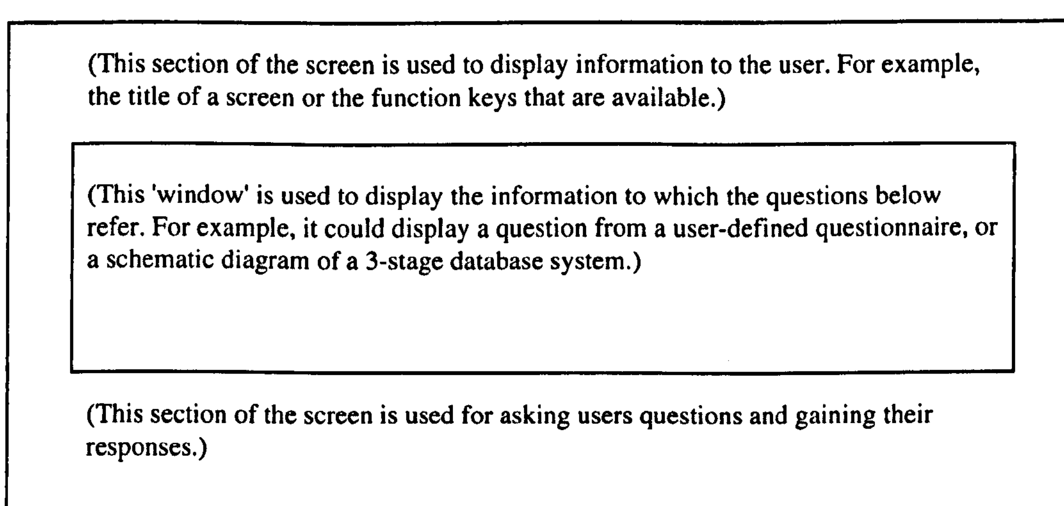


Figure 4.4: Basic layout of the tools' screens

As the tools will be used by non-computer experts, their interfaces should not contain specialist computer terminology. Instead, they should rely upon general terms commonly associated with a construction task, especially as the tools will only be used infrequently. Hence, for questions with a limited list of possible responses, menus would be the best method, thereby enabling the available options to be displayed to the user. For a sequence of related questions, a form filling structure should be adopted.

The goal of the suite of tools is to construct an appropriate IDDA for an expert. This overall goal can however be broken down into various sub-tasks and separate tools built to achieve each one. Thus, the tools lead an expert through this series of tasks, thereby ensuring that all relevant details are gathered. This approach will allow a user to concentrate on one small section of the overall problem at a time. This should reduce the possibility of overloading the user and at the same time both appearing to follow a very structured approach during the development of the IDDA, and appearing to provide extra assistance to the inexperienced system builders.

For this research project, four sub-goals were defined for the tools:

a) to develop appropriate databases for the domain,
b) to develop appropriate means for collecting and storing the data,
c) to include any domain specific help,
d) to develop appropriate means for reviewing and analysing the stored data.

Therefore to achieve these goals, a variety of information sources are required. As already explained, the assessment questionnaires will be the main source. These will have been devised by the domain expert and word-processed into ASCII text files by the departmental secretaries (see Appendix C (I)). The domain assistant then interacts with the tools to provide the final pieces of information before the tools automatically construct the IDDA system. This ability to spread the quantity of work involved in building an IDDA system around the different groups of people within the specialist domain, enables the various talents of the groups to be harnessed. Moreover, this distribution of the work amongst the domain team will also reduce the demands placed on each participant.

The details required from the domain assistant relate to each of the questions in the assessment questionnaires, e.g. the answer type, length, range and when the question is asked. These details will allow the tools to build an appropriate IDDA system with no further interaction with a user during the construction phase. For the data entry interfaces of the IDDA system, appropriate question and answer definitions as well as the question flow must be specified. Since the information to be stored in the databases will correspond to the answers to the assessment questionnaires, a user need only define the type of response expected and the actual questionnaire navigation.

There exist four different categories of response: single answers, a column of answers, a table of answers, and, comments (as shown in figure 4.5). Each category other than comments can be of the type: numeric, logical, alphanumeric or a date field. A table can consist of a number of different types of response. However, a table will have either its columns and/or its rows containing the same response type, depending on its format.

F6 - ABANDON   F1 - Help
Designing the Assessment Form   F3 - QUIT&SAVE

Question Number :-   3

Gender:
1 = Male
2 = Female

Will the answer be :   [ ]

Numeric (Enter N)
Alphanumeric (Enter A)
Logical e.g. Y/N, T/F (Enter L)
Date (Enter D)
Table (Enter T)
liSt (Enter S)
Comment (Enter C)

Figure 4.5: The options available for questionnaire answers

As explained in Appendix C (I), the tools attempt to infer the response type for a question from the ASCII question file. If the type entered by the user does not correspond to the inferred type, the tools request that the user confirms the entry. Mistakes can therefore be quickly detected and corrected, thus providing assistance for the naive computer users.

Further information needs to be acquired concerning the answers to the questions (see figure 4.6 for an example of a numeric question). If a field is:

a) logical - is it Yes/No or True/False?
b) alphanumeric - what is the length?
c) numeric - what is the maximum, minimum and the number of decimal places?
d) comment - is it attached to the previous question or the following one?

F6 - ABANDON   F1 - Help
ESC - RESPECIFY   Designing the Assessment Form   F3 - QUIT&SAVE

Question Number :-   3

Gender:
1 = Male
2 = Female

Numeric.

Enter the number of decimal places (enter 0 if none) : [ ]
Enter the maximum possible value of the answer : [ ]
Enter the minimum possible value of the answer : [ ]
Enter the most common value for the answer : [ ]

Figure 4.6: An example of the responses required for a numeric questionnaire question

In addition, the most common response to the question needs to be defined. The value specified as the most common response will appear in the answer box for the question when the data entry interfaces of the end-system are displayed. The user entering information into the IDDA system therefore needs only press the return key for the displayed response to be accepted. If the answer is different to this default, the user merely enters the appropriate reply. The literature shows that naive computer users prefer interfaces with default values, especially if they are intermittent users of the application (see Section 4.2.2). If the domain expert has carefully considered the most appropriate default responses, their inclusion should reduce the amount of time and effort required by the user of the IDDA system during data entry.

With assessment questionnaires, form navigation is reliant upon answers to previous questions. Normally navigation, in either a manual of a computerised scenario, is not dependent on more than two previous answers since an inquirer cannot be expected to follow a complex format and interview or assess a person at the same time. In general, it is only dependent upon the response to the previous question. To enable form navigation to be mapped in an IDDA system, appropriate details must be defined. This is achieved by the specification of when each question in the questionnaire is to be asked, i.e. either 'always' or 'only when ---'. If the selection is 'only when ----', the relevant previous questions and their responses must be defined. For example (see figure 4.7), a question asking 'Date of Re-injury' should be asked 'only when ---' the previous question, 'Re-injury - Y/N ?' was answered 'Y'. The tools check that the condition specified by the user is valid, i.e.:

- that an earlier question number to the current one has been selected,
- that the comparator sign is appropriate (e.g. for a logical question, only <> or = can be selected and therefore only these options are shown),
- that the response entered is of the correct type and is valid (e.g. for a logical question, the answer response for the question will have been classified as either a T / F or Y / N category and therefore a valid entry here will depend on which category was specified).

F6 - ABANDON   F1 - Help
ESC - RESPECIFY   Designing the Assessment Form   F3 - QUIT&SAVE

Question Number :-   8

Date of Re-injury :

Enter the question number involved in the condition :   [7]

Select - 1/ EQUALS (=)   2/ NOT EQUAL (<>)   [2]

Enter the comparator : [Y]

Figure 4.7: An example of attaching a condition to a questionnaire question

Multiple conditions can be defined, linked together by ANDs or ORs, and brackets can be inserted, if required. This information will therefore provide the tools with all the relevant details to determine form navigation and hence program control.

Three text files for each assessment stage are created by the tools to store this information (see figure 4.8). One contains the variable details from which the appropriate database for the stage will be constructed. Another holds the questions, now separated and with additional information on the expected answers. The final text file stores the description of the variables and the time when the questions are to be asked.

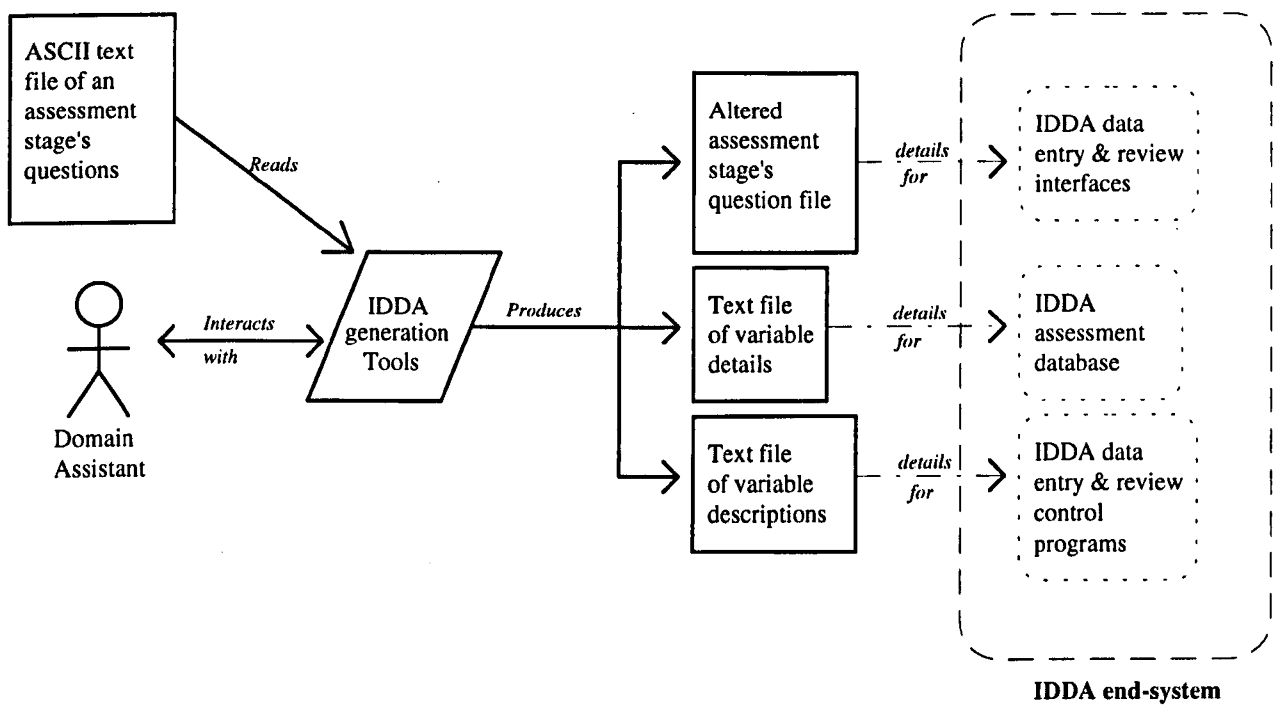


Figure 4.8: The inputs and outputs of the Tools

Therefore, from this information the tools can:

a) build the databases to store the details gathered at each assessment stage,
b) develop the interface programs for data entry,
c) construct the programs for viewing the entered records,
d) develop the programs to control the data entry and the review of records.

The interfaces for an IDDA end-system follow similar design criteria as those used in the design of the interfaces for the tools (for an example, see figure 4.9). They are based on menu and form-filling styles. The background colour for all screens is dark green. Yellow is used for the foreground colour except for error messages and function key labels, which are white. Any answer boxes appear in a dark red-brown colour and users' responses are in white. Error messages are displayed on the bottom line of the screen whilst the function key labels always appear on the top lines. Menus are centred and questions involved with form-filling interfaces are left-justified. The size of an answer box is determined by the maximum possible length of the answer to a question. In any data entry screens or review screens, their positioning is right-justified but is also dependent upon question length and answer length. During data entry or review, the function key labels, the title and the first two questions of a questionnaire, i.e. the reference questions (in figure 4.9, these are 'STUDY NUMBER' and 'DATE OF ASSESSEMENT'), are always displayed. Therefore a user is continually aware of which record is being reviewed, amended or added to which questionnaire database, as well as which functions are available and how to initiate a desired function.

```
 F2 - HELP                                   F6  - Exit/NO SAVE
 F3 - QUIT & SAVE    Initial Assessment      F10 - PAGE FORWARD
                     ------------------      F9  - NO VALUE

Study Number                                                  1

Date of Assessment                                   16/10/1990

Gender             1. Male                                    1
                   2. Female

Date of birth                                        28/11/1960

Date of injury                                       28/11/1989

Injured knee       1. Right                                   1
                   2. Left

Other knee normal                                             Y

Re-injury?                                                    N
```

Figure 4.9: An example of a data entry screen from an IDDA end-system

The interfaces of both the tools and IDDA end-system are therefore designed to be natural, consistent and supportive for users as well as minimising redundancy of information presented and requested (see Chapter 3 for justifications). In any system, help is a very important facility especially for novice users. The next section describes the various help facilities provided by the tools and IDDA system.

4.5 HELP facilities

Studies on help have generally been based on the assumption that users can be classified into two categories, novice and expert computer-users. In this research, the users of the IDDA system and tools will be novice computer users, though they may well be experts within the specialist domains in which the generated IDDA end-systems are to operate. This section therefore deals primarily with a discussion of the provision of help for novice computer-users.

Novice computer-users need help frequently and may in fact spend most of their time searching for information (Vaubel and Getty, 1990). However, as they point out, they will not use a help system they cannot understand. A failure to match the comprehensibility of help information to the expertise of the user can lead to unfortunate results. Draper (1984) found in his study of UNIX command usage that the use of on-line help facilities increased with expertise. Draper (1984) believed that this was mainly due to novices not knowing how to access the help and also because novices could not comprehend the help once it was displayed.

With humans, the help and explanations given by an expert depends on an assessment of what the questioner has failed to understand and the knowledge of how much the questioner knows about the domain. In this research there are two systems, and although in both cases the users may well be novice computer-users, their expertise of the tasks to be accomplished in the two scenarios is different. Therefore the type of help required in the two situations must be different.

With the tools, users require to know how to operate the tools and what information the tools expect. Tasks faced by users are construction tasks, i.e. the design and development of IDDA systems. This will generally be alien to a user. Help is therefore required to explain the stages involved in the construction process and how the questions being asked by the tools should be answered.

With the generated IDDA, the users will be familiar with the specialist domain. They will understand the terminology used in the questions though may not know how to acquire the answers, i.e. how to carry out a particular test or how to take a specific measurement. The help therefore is to remind users of procedures they may have forgotten.

In this second scenario, the best person to determine the help required, and when help is likely to be requested by a user, is a person who is an expert within the specialist domain, e.g. the person who designed the questionnaires. Consequently, a tool is required to enable user-defined help to be specified and linked into the IDDA end-system. With the first scenario, however, the best help will be produced by the computer specialist who built the tools and who can guide users through the construction tasks and attempt to anticipate the problem areas.

Mager (1983) conducted an experiment with novice users on the variable content of help and error messages. They modified an old system to supplement the help command with a help key and made the content of the help and the error messages more "concrete". They discovered that those who used the modified system completed typical office tasks in less time and with more positive attitudes than those who used the original version of the system. This seems to indicate that the allocation of a standard method of access, i.e. a specific help key, and the presentation of context sensitive help influences not only the speed of operation but also how a user perceives the end-system. Furthermore, Borenstein (1985) found, from his study of five different help systems, that it was the quality of the help text rather than the attempt to provide sophisticated access methods that was the most important factor in determining the usefulness of a help system. Therefore the evidence seems to indicate that a simple standard method of initiating help is all that a user requires and it is, in fact, the quantity and quality of the help presented that are influential factors.

In its earliest form, the text of on-line help was generally the same as the text of the hard copy documentation. This text was displayed on the screen when a user requested help. The text occupied the entire screen and replaced the entry screen that a user had been viewing. To resolve this problem, the help text was then presented in split screen or windowed format. This enabled a section of the entry screen to remain visible to the user while the help text was being displayed.

The major benefit of full-screen help is that it takes advantage of the entire screen area and therefore this allows a large amount of help text to be shown on each screen. However, users cannot see the entry screen and the help text at the same time. If a specific instruction is needed, users must either remember it until the display reappears and they complete the task, or they must write the instruction down whilst consulting the help facility.

With a split screen format, help text blocks 50% of the screen and although parts of the initial entry panel may be accessible, some fields are covered by the help text. Furthermore, in contrast to the full screen format, a split screen can display only half as much help text.

In a windowed environment, the help panel rarely occupies more than 30% of the screen and, therefore of the three help formats, this one displays the minimum quantity of help. A further problem, with both the windowed help and the split screen, is in ensuring that the help does not cover the section of entry screen to which it refers. This implies therefore that the help cannot be displayed in a fixed position. Berry and Broadbent (1987) found that this caused problems, 'multiple windows popping up and disappearing all over the screen can be very disorientating especially for naive users.'

Cherry et al (1989) carried out a study of the three methods of presenting help - full-screen, split-screen and the window technique. They recorded the views of the users. Some users said that they liked the split screen and window approaches because they felt that the task was not being interrupted. However, many reported that they did not like scrolling through the help windows. In addition, the users complained that when using the windowed format there was sometimes a confusing break in the text in the window. Windowed headings occurred frequently and because of the minimal area available, text on a panel was often limited to one or two sentences. The help was therefore difficult to read and appeared fragmented.

With a full-screen help, far more text appears on one screen thereby reducing the need to scroll and in some cases eliminating it completely. Furthermore, the increase in space enables the layout and continuity of the help to be reviewed by the designer. However, as the help covered the original display, the users found it difficult to relate the help back to the problems they were having on the entry screen (Cherry et al, 1989). Consequently, there seems to be no simple solution to the question of which is the optimum format.

Hypertext is another approach. However, it too has problems. The two most commonly quoted are disorientation and cognitive overload (Conklin, 1988). These occur as users lose themselves in a web of links due to the extra feedom given to them to move around the text. Moreover, it seems that hypertext is best suited to a fairly large interrelated task or explanation, e.g. the exploration of an encyclopaedia. Smaller, highly directed tasks seem better suited to context specific help (Wasson and Akselsen, 1992).

Generally, when users request help, they need clarification on the question being asked or the type of response that is expected. As the evidence above shows, if a help screen covers the question and answer slots concerned, users have the difficulty of trying to link the displayed advice or instructions to their specific problem. To try to adopt a split screen or windowed approach can cause even more problems. Consequently for this research project, it was decided that the best approach would be to present the help as an electronic text book - as something for the user to refer to when dealing with a problem but in a way that does not interfere with the original display.

This could be achieved by using a second monitor. The main monitor displays the original interface where all entries are made, but any help that is requested appears on the adjacent monitor. No interference therefore occurs to the original layout and users can refer to the second screen for advice and examples to assist in overcoming any difficulties they have. Having a totally new monitor on which to display help, enables a full screen of text to be presented. This reduces the need for scrolling and enables fragmentation of the text to be minimised. Consequently, readability and comprehensibility should be enhanced whilst at the same time reducing user annoyance and frustration.

In addition, unlike hard-copy, an on-line help system ensures that it can neither be 'lost' nor damaged, therefore, it will always be available when required. It can be programmed to be context-sensitive so that the explanations displayed are relevant to the problem confronting the user at that moment in time. This will stop the user having to search for the correct answer through a mass of information, thus reducing cognitive overload and preventing user disorientation and time delays. With the second monitor positioned next to the main monitor, help displayed on the second monitor is easily seen by users as it is inside their natural line of vision. Therefore it is easy for the user to cross-reference the information on the second monitor whilst answering a question on the main monitor. A final point regarding the use of a second screen is that the purchase of a dumb terminal is still relatively cheap when considering the cost of large, high resolution screens. Furthermore, many hospitals have unused dumb terminals left over from the implementation of old information systems, hence in some situations purchasing new dumb terminals may not actually be required. It was, therefore, decided that a second monitor would be used for displaying context-sensitive help during this project and that this configuration would be reviewed during the evaluations.

With this project, there are in fact two distinct systems, each requiring specific help. One is to be associated with the design and the construction of an IDDA system whilst the other is to offer assistance when the finished product is running. Both are described below.

4.5.1 System HELP

The system HELP, activated by using F1, will be available to users during the development of an IDDA system, e.g. when using the tools. It will always be available to users and will be written to be relevant to the particular problem at hand. It will also include examples where appropriate.

4.5.2 User-defined HELP

The user of the constructed IDDA will also require help. However, the help required will be associated with the specialist domain in which the IDDA end-system will operate. Therefore, as the IDDA has been built by domain experts, the only way of ensuring that the help is relevant to the problem is to allow domain experts to define the help themselves. Not only are domain experts the best qualified to specify the help required, but, in addition, they will use the specialist terminology of the domain and will include relevant examples.

The procedure for developing user defined help will involve the following steps. The expert first defines, on paper, the help for each of the stages. This information is then entered into a word-processor and ASCII text files are produced. The secretaries are again the best qualified to carry out this task. Once the ASCII files have been created, the domain assistant interacts with the tools to link the various sections of help to the corresponding questions in the appropriate stage assessment.

The end-users of the IDDA can consequently obtain relevant help to assist them with any problems that may arise in completing the various questionnaires. This help is displayed on the second monitor in the same manner as the system help. If no help has been defined by the domain expert for the question, an appropriate message appears on the second monitor to explain this.

4.6 Other facilities

Other facilities that are available to users of the tools are:

a) the ability to ABANDON the whole construction process and start again redefining an IDDA system,
b) the ability to SAVE everything entered AND RESUME from this point onwards at a later date,
c) the ability to SCROLL the current assessment question being defined on the main monitor, where appropriate,
d) the ability to GO BACK and REDEFINE the answers describing a particular question.

During the data entry phase, in the actual IDDA system, the user has the abilities to:

a) PAGE FORWARD onto the next entry screen,
b) QUIT and SAVE the current record even if it is incomplete,
c) Enter NO VALUE as a response to a question. These values are not included in any analyses,
d) EXIT without saving the current record or any changes made to it.

The facilities offered have been designed to aid acceptability and to put the users at ease when they are operating the tools or the IDDA system, even if they are naive computer users.

4.7 Summary of the stages involved in the construction of an IDDA system

Figure 4.10 illustrates the stages involved in building an IDDA system and the various interactions that take place. These processes include:

a) establishing an initial overview of the system to be built,
b) obtaining the details to construct the required databases,
c) obtaining the details to develop appropriate interfaces and methods to acquire the data,
d) building the databases,
e) linking user-defined HELP to the appropriate questions,
f) building both the interface and the control programs for data entry and storage,
g) building the appropriate review programs.

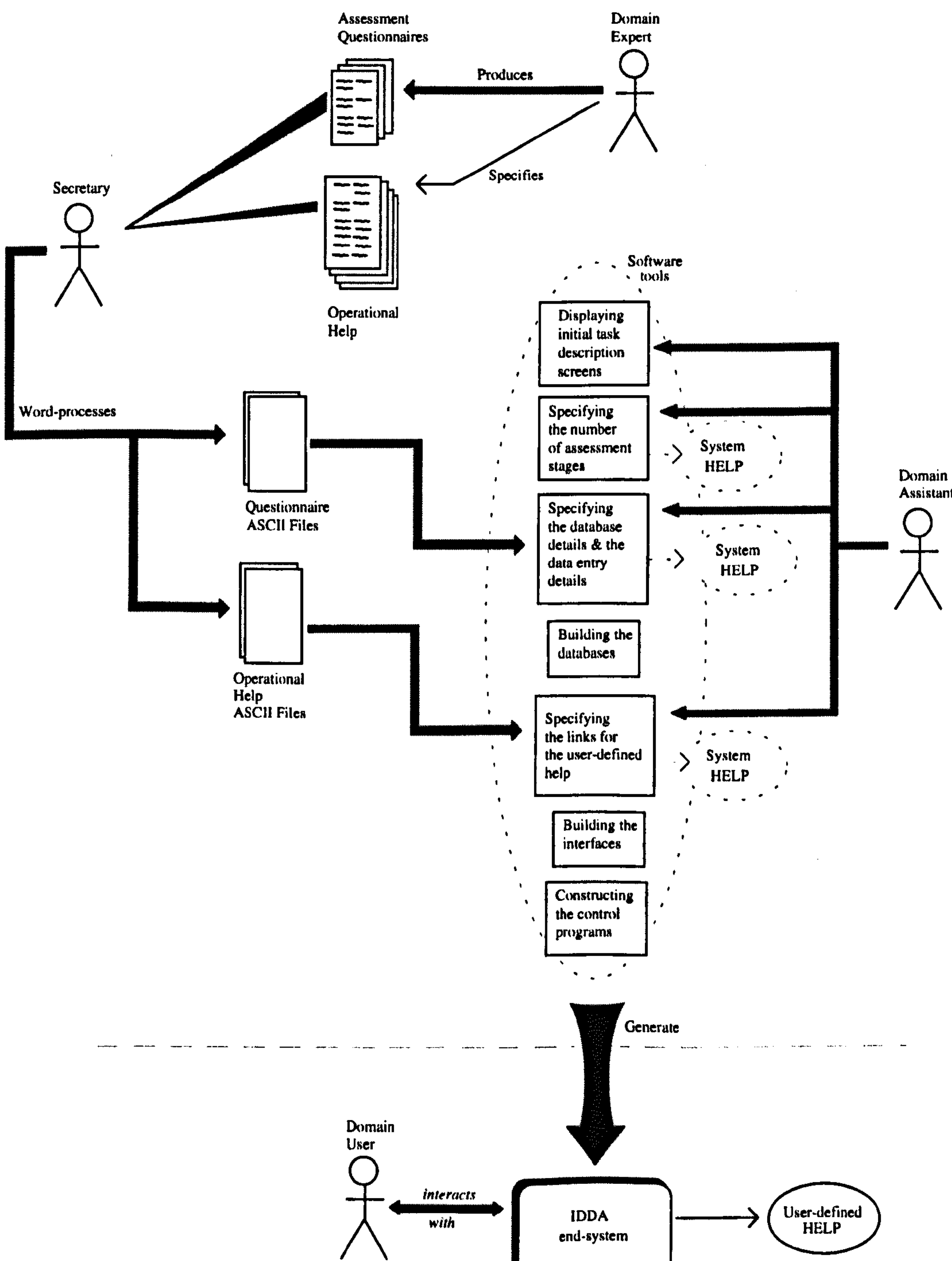


Figure 4.10: Stages in building an IDDA system

Chapter 5

Problem-solving and Decision-making

5.1 Introduction

The processes by which humans routinely reach decisions are still not completely understood. As Brooks (1991) explains, 'the way in which our brain works is quite hidden from us. We have some introspection, we believe, to some aspects of our thought processes, but there are certainly perceptual and motor areas that we are quite confident we have no access to'. However, the goal of expert systems has been to replicate the responses of a human expert during decision-making by following decision pathways and mechanisms perceived to be similar to those used by a human expert. Thus, in theory, the system should reach the same decision as the human. Unfortunately, it is evident from the limited success of these systems that this goal is as yet unattainable. The few recorded successes have been in situations where the decisions, which were automated, were in fact 'mechanical' for the human in the first place. Hence, as Lings et al (1991) warns, 'the danger is in persuading ourselves that, because computer programs have been shown to replace a human when it comes to simple, mechanical decisions, it therefore follows that all decision-making by the human within the organisation can be similarly replaced'.

Even AI supporters disagree over the depth and nature of intelligence which can be adequately and accurately represented in a computer system (Andriole, 1985; Brooks, 1991). One of the fundamental problems is in attempting to define intelligence itself, never mind endeavouring to model it. Moreover, there is the realisation that any 'smart' system, though narrow in its focus, must degrade gracefully at the edges of its expertise rather than being as Andriole (1985) describes, 'brilliant in one area and completely unintelligent in another'.

Thus there is an emerging perception of AI systems as merely being extremely limited extensions of relatively mundane human information processing activities whilst others suggest that it is time to reconsider current AI objectives, 'we sometimes need to step back and question why we are proceeding in the direction we are going and look around for other promising directions' (Brooks, 1991). However, these views have only recently started to gain support as they differ quite considerably from the orginal perception of the capability of AI systems.

The following sections briefly outline human problem-solving and medical decision-making. A number of the difficulties inherent in modelling these processes is highlighted before the possibility of utilising statistics is explored. Finally, there is a description of the facilities which have been incorporated into the IDDA end-system to enable it to provide assistance to human experts investigating their specialist fields.

5.2 Problem-solving

A problem space is composed of a description of the initial state and the goal state, a set of operators to transform one state into another, and knowledge of search control which guides the selection of which operator to apply or which state to treat as the current state (Guidon, 1990). For example, checking the logical validity of an argument is a problem with a clearly defined goal for which the subject must find an appropriate method of solution. Similarly, every decision to be made is effectively a problem to be solved.

In fact, it is the ability to think in a highly selective manner that enables humans to function in the real-world and is one reason why computers have not yet succeeded in these situations. 'We select from the vast amount of information available from sensory input and memory storage, the relevant information needed to solve a problem. We formulate and choose a few of a vast number of possible operations that could be applied. It is in our ability to select successfully that we are vastly superior in intelligence to any computer that has yet been built' (Evans, 1983).

There is a need therefore to have a clearer understanding of the decision-making process of a human expert. Jacob et al (1986) believes that the type of work needed here is experimental, where studies can be initiated and the results analysed. Parsaye and Chignell (1993) agree, 'information is usually obtained by performing experiments, by stumbling upon discoveries, or by getting the information from someone who already has it. In today's information society, information is more valuable than ever before, and its value continues to increase'.

Understanding, prediction and learning all require information. In fact, the progress of society is dependent upon the continual efforts to extract the lessons of the past into knowledge which can be used for making better decisions in the future. This process requires the collection of data and examples to enable analyses to be undertaken, thereby uncovering new data patterns and confirming previous observations and beliefs. In this manner, models of the world evolve, providing insights and explanations for the relationships in and between data.

Humans are generally very good at identifying patterns or differences within a data set although they are poor at accurately recalling, analysing and reviewing past cases to validate a theory. The vast amount of available information is forcing investigators to seek assistance. Computers are superbly equipped for providing such support. However, the initial decision-support systems that were built attempted to emulate the human expert rather than work alongside in a co-operative partnership. There was not therefore the ability to enable the discovery of 'knowledge from a database that may then be pooled with human knowledge and expertise in information interpretation and decision-making' (Parsaye and Chignell, 1993). Consequently, the benefit of combining the strengths of both, the machine and the human expert, were lost.

One reason for adopting a joint approach is the existence of variance in problem-solving strategies among experts, which makes the building of computer systems to mimic human problem-solving fraught with difficulties. Decision-making is based on intuition, past experience, preferences and different problem-solving skills and involves a variety of techniques such as analogies, lateral thinking etc. Davis (1989a) defines intuition as 'the ability [to] effortlessly and rapidly associate with one's present situation an action or decision which experience has shown to be appropriate'. However it is extremely difficult to describe and is specific to an individual, making it practically impossible to formalise and model within a computer system. Consequently, as Davis (1989a) states, 'in no sense can one capture human expertise and store it in the form of a complex reasoning system if human expertise is an intuitive associative ability not based on processing facts by means of rules'.

Moreover, contexts vary since generally each expert has a different environment of problems. It is therefore not surprising that two experts in a domain often have different sets of rules-of-thumb, thus leading to the consensus problems mentioned in Chapter 2. These differences have been so noticeable that Hand (1985) suggested that there seemed to be a different problem-solving style for each clinician, which was only constrained by the common underlying process of generating and testing hypotheses. Although, Hand (1985) claims that this leaves greater flexibility in designing programs for diagnosis, it creates major acceptability problems for a developed system when the clinicians are so different.

One of the sources of these differences is the cognitive limitation that effects human problem-solving performance. As explained in Chapter 2, humans have a large capacity long-term memory for storing facts, principles, events, and knowledge of various sorts, but a severely limited short-term memory. The capacity of short-term memory is reputed to be only a few (between four and seven) familiar symbols or chunks. These chunks are packets of integrated knowledge about the domain, both procedural and declarative. One or more of several kinds of knowledge can be contained in these chunks, e.g. procedures, heuristics, information that guides selective perception and pattern recognition, and information that allows selective retrieval. Their role is therefore to aid problem-solving. In fact, each chunk can often be applied to more than one problem (Garg-Janardan and Salvendy, 1988).

It is claimed that the limitations of short-term memory prevent problem-solvers backtracking from an unsuccessful solution path, as the previous steps would have to be stored. Instead, humans tend to focus almost exclusively on proceeding from the current situation, whatever that may be, with any consequences associated with such a manoeuvre (Simon, 1990).

Evidence for these theories was gained from investigating problem-solving in highly structured tasks, i.e. playing chess, proving theorems in logic etc.. These experiments revealed the several characteristics of the human problem solver stated above - a limited capacity of short-term memory, a use of heuristic strategies to examine promising avenues, a tendency to search for information sequentially, and the importance of the problem solver's conceptualisation of the problem at hand (Newell and Simon, 1972).

Researchers studying the strategies involved in playing a game of chess began to believe that a successful model of this problem space, e.g. a program which could beat a grandmaster, must possess and be exhibiting the same problem-solving strategies as those used by the human expert. However, the successes recorded in chess playing programs have in fact been driven by advances in technology, which have enabled larger and quicker searches to be undertaken. As Brooks (1991) points out, 'for any given technology level, a long term freeze would soon show that programs relying on search had very serious problems, especially if there was a desire to situate them in a dynamic world'. In reality, therefore, current chess programs are not very good models for general human thought processes.

Brooks (1991) reported on further evidence supporting this view. Here the game of Go was involved. In this game the search tree is much larger than for chess and a good static evaluation function is much harder to define. Go has never worked well as a vehicle for research in computer games. Brooks (1991) believes that, 'any reasonable crack at it is much more likely to require techniques much closer to those of human thought - mere computer technology advances are not going to bring the minmax approach close to success in this domain'.

However, for both of these games, Go and chess, the problem space and the rules have already been defined. In many situations in the 'real world', there are no or few such specifications. Therefore, the problem-solver must first attempt to understand the problem that is being presented (Simon, 1990). In the case of familiar problems that a problem-solver has encountered previously, the understanding processes are determined by those previous experiences, which may be different for different subjects. Schank (1990) believes that more-effective understanding manifests itself in better predictions about what will happen in particular well-constructed experiences, which have been built up over time. However, Schank (1990) warns that these predictions are only as good as the initial categorisations of the world made by the subject.

Thus, the view is that problem-solving employs two complex processes: an understanding process that generates a problem space from the text of the problem and a solving process that explores the problem space to try to solve the problem. In humans these steps do not proceed exclusively to completion (Simon, 1990). Instead there is frequent switching between the understanding process and the solving process. The solving process appears to exercise overall control, in the sense that it begins to run as soon as enough information has been generated about the problem to permit it to do anything. When it runs out of things to do, it calls the understanding process back to generate more specifications for the problem space. As recognition of particular features in the situation evokes new elements from long-term memory, the solver's problem space undergoes gradual and steady alteration (Simon, 1990).

5.3 Medical decisions

The major focus in medical decision-making has been mainly concerned with diagnosis.

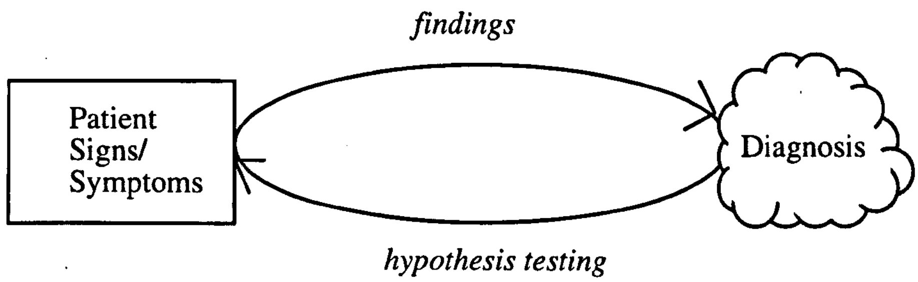


Figure 5.1: Medical diagnosis

This involves the determination of the initial states, which is the first stage in problem-solving. Medical diagnosis is an iterative process (see figure 5.1) and has been described as consisting of the following steps:

'1) While the physician is looking at a patient, he collects signs. These signs evoke (let the physician think of) a set of pathologies *(sic)*. Some of these are discarded because they are strictly incompatible with the current case.

2) The remaining hypotheses are compared and the physician tries to apply differential diagnosis. The physician then evaluates the hypotheses and isolates those which appear most likely.

3) If one or more hypotheses are confirmed with sufficient certainty, the physician states his diagnosis. Otherwise, he goes back to point 1) and tries to obtain new signs allowing a differential diagnosis between the remaining hypotheses' (Du Bois et al, 1989).

Studies of the behaviour of expert physicians over the past several years at the University of Minnesota and elsewhere suggest that authentic reasoning in medical diagnosis is based on a blend of symptom-centred and disease-centred knowledge (Johnson, 1983; Kunz, 1984; Reggia and Tuhrim, 1985). In its idealised form, such reasoning consists of hypothesis-testing with an initial symptom-centred phase in which the diseases hypotheses are generated on the basis of carefully chosen patient data. Because there are innumerable 'facts' which could be gathered, there is a need for a sharp focus for this activity. This focus is obtained through the pursuit of a small set of diagnostic hypotheses that are suggested by the presenting complaints. Disease-centred knowledge associated with these active hypotheses then provides expectations for clinical findings. Hence a comparison is made between the pattern of findings in the patient and the consultant's concept of various features of diseases including known clinical characteristics, the evolutionary changes in clinical features, the predisposing factors, and complications. Such knowledge guides the collection of additional data which, once collected, can then be compared. The evaluation of a hypothesis ultimately requires the assessment of how many findings caused by a given disease are present and how many expected findings of a given disease are absent, i.e. a judgement is made regarding the degree of 'fit' (Pauker et al, 1976; Kassirer and Gorry, 1978).

The majority of current medical expert systems have attempted to model this process and thereby give assistance to physicians during the diagnostic process. Nevertheless, once the initial state has been established, i.e. the identification of the problem or the disease at hand, a decision still needs to be made regarding the action or treatment strategy to be taken. Current medical expert systems however rarely venture into this area of the decision-making process. One reason could be that the highly specialised medical domains selected for these systems have clearly defined treatment paths once the problems have been identified and defined.

There are situations, though, when this is not the case. The findings might be such that the diagnosis of the problem is relatively straightforward, i.e. a broken bone. The uncertainty, or problem, occurs when deciding on which treatment strategy to adopt (see figure 5.2).

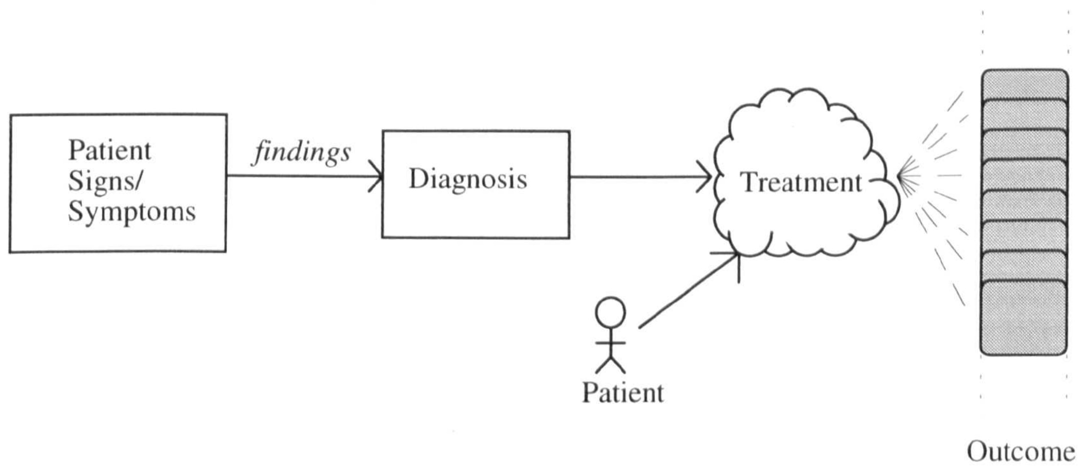


Figure 5.2: Treatment selection

The physician, again, gathers the evidence on either side of the treatment or outcome question, attaches weight to it and rules on its admissibility. The physician then attaches numerical probabilities to these judgements according to the current situation. These probabilities are not the result of an analysis of real-life data but are the physician's own subjective probabilities of events and are based on an individual perception of the situation, i.e. previous similar experiences that can be recalled from memory.

Physicians often use such probabilities during discussions with other colleagues and during decision-making. As Gage (1993a) describes, 'physicians discuss medical issues in terms of probability; 'These electrocardiographic changes are associated with myocardial infarction 50% of the time.' 'The chance that the unit of blood contains HIV virons is 1 in 250,000'. Such expression are familiar to us'. However, there is no demand by either peers or by administrative authoratives for physicians to record and justify these measures with actual real-life data.

The concept of subjective probability is based on the premise that everyone has a degree of belief concerning the likely occurrence of some event relevant to them and their environment. The value of the concept is that it allows decision-makers to describe their feelings about the effects of uncertainty in defined and understood numerical terms, and thus to incorporate their judgments explicitly into the decision process. As Moore and Thomas (1976) state, 'the resulting numbers do not imply objectivity or authority; they are an understood way of putting subjective views into a more precise form providing, in turn, a basis for comparison when relating one decision-maker's evaluation with another.' Thus, as judgments differ, so may subjective assessments. This is supported by Tversky and Kahnemann (1990), 'in reality, subjective probabilities determine preferences among bets and are not derived from them'.

Therefore, the subjective aspects of a physician's thinking process have not been clearly identified. Researchers have continued to demonstrate the variability in the strategies adopted by one physician as opposed to another (Reggia and Tuhrim, 1985). Consequently, the mechanisms used to reach a decision, and in fact the decision itself, are highly likely to be different for different physicians - 'the differences in their methods for reaching consensus are striking. Indeed, as we have said, the medical profession does not even appear to have a definable method' (Gage, 1993a). Hence it is often very difficult to ascertain the relationship between the scientific evidence and the medical decision reached.

This leads to great uncertainty. As Gage (1992b) explains, 'the uncertainty we feel about those whose performance we are judging reflects our own uncertainty about how to perform the tasks they are attempting. Logically, it cannot be otherwise. If we knew the rules, we would know whether they were being followed. In medical training, we try to judge the ability of our trainees to make the right clinical decision under conditions of uncertainty. Unfortunately we do not always know what the right decisions are'.

Gage (1993a) believes that subjective elements within the decision-making process are the major cause of the uncertainty. Whereas the objective measure can be represented in a numerical probability, the subjective judgement is harder to define. For example, in tossing a coin, there is the subjective judgement that the coin is a fair coin which enables the objective probability of 0.5 for gaining a head or a tail to be made. Both are important and must be present when a decision is to be reached. Therefore, Gage (1993a) suggests that the subjective elements should also be quantified in terms of probabilities that have been examined. This can be achieved by recording the assumptions initially and then later analysing them with the known outcomes and objective measures. Hence as Gage (1993a) explains, 'we may see the possibility for the use of true probability distributions, albeit subjective probability distributions, in medicine. Then, consensus can be articulated in quantitative terms'.

Experience is one such subjective element that does play an influential role in problem-solving. It is used as a mechanism to update a person's knowledge-base, i.e. successes reinforce known rules or hypotheses, failures cause a re-analysis of previous reasoning strategies. Thus, previous cases or experiences often act as examples in future decisions. In medicine, decision-making is often based on this type of associative knowledge. 'Medicine is a field in which we use our own personal experience to guide most of our decisions. And as any expert in expert systems will tell you, these decisions can rarely be reduced to a set of rules' (Gage, 1992b).

However, medicine has a distinct advantage over many other domains as documentation generally exists for most of the cases. Consequently there is a large quantity of collected information describing the progress of individual cases, e.g. the actual experiences which occurred in particular situations. This data could therefore be analysed to help alleviate uncertainty and thereby assist in future decision-making. As Gage (1992b) explains, 'we should be aware that to reduce the uncertainty that will confront us as we make clinical decisions in the future, we must first all reduce our uncertainty about the decisions we have already made. Accurate, detailed record keeping will help us reduce our uncertainty and improve our ability to train future physicians'. Unfortunately, as Gierl and Stengel-Rutkowski (1994) points out, 'the intrinsic medical experience is not yet used in knowledge-based systems. On the contrary, even inductive rule-based systems comprise of only the subjective parts of medical experience hidden in every case a physician encounters'.

This situation must however change with the continuing fast growth in medical knowledge. Physicians are now perpetually being forced to specialise further, thereby making them reliant on other experts when they are presented with problems outside their expertise (Shortliffe et al, 1984). Unfortunately, the techniques used by humans to relate, synthesise and apply information are still vastly superior to the methods used by computers. However, for the detailed recall of a particular patient's characteristics, diagnosis, medication, outcome, etc., a computer database far exceeds the accuracy obtainable by a human, especially when dealing with numerous cases over a number of years. Consequently, database systems would be capable of, 'increasing the quality of activities that are dependent on vast amounts of information by helping people to better manage that information (i.e. dealing with the so-called information explosion)' (Blascovich, 1987).

Therefore, rather than relying solely on the subjective values of individual consultants, the emphasis of this research has been to enable values to be arrived at statistically by examining previous case histories. In this manner, the elements that effect the outcome can be ascertained as can the level of influence they will exert and the likelihood of a favourable outcome. This knowledge could therefore assist the consultant in selecting the most appropriate treatment for each particular patient. Consequently, instead of resorting to 'gut feeling' decisions, there will be a more rigourous and justifiable method available.

In the past however, there has been great interest in developing medical expert systems, though they have not managed to achieve the success expected of the technology. Many factors have been attributed to the cause of the problems, for example:

- there was no effort to fit the system to the environment or to the special needs of the medical departments, such as the entry of patient data, management of patient data and others,
- the ambitious inference process and knowledge representation far exceeded users' understanding and background,
- users could not handle the demands of continually having to update the knowledge base, which was the requirement of the knowledge acquisition component of the system (Gierl and Stengel-Rutkowski, 1994).

Thus, it seems a more appropriate system would be one that:

- requests only the essential details to be entered with the minimum amount of additional interaction,
- is straightforward and easy for users to learn, use and understand,
- provides efficient and effective support in the daily activities of its users,
- requires little maintenance or updating.

As Gierl and Stengel-Rutkowski (1994) state, 'the problem, often overlooked by data processing professionals, is that medical activities are much more complicated, vague and information-intensive than those in an industrial setting. Therefore 'simple' things like reports, drug orders, duty rosters, etc. are in many cases, the only domains which could be supported by expert systems up to now'. Consequently, the indications are that only well-defined, well-structured tasks are capable of being formalised and modelled with current technology. Any task that can not be defined by rules, e.g. those requiring intuition, is not suitable for such systems. Therefore the initial problem space must be analysed and reviewed very carefully to determine whether it can be modelled appropriately and if so, which technique is best suited to the task.

5.4 Well-structured and ill-structured problems

It has been suggested that a well-structured problem (WSP) is one which can be fairly easily mapped into a representation which permits formal or mechanical problem-solving techniques to be applied. Thus, for example, the game of noughts and crosses is a WSP. The problem solver has a clear knowledge of the goals, constraints and methods available. These problems are usually of a kind where an initial state can be changed into a goal state by a series of transformations (Hand, 1985; Evans, 1989).

Many real life problems, however, are ill-defined or ill-structured in that neither the goals nor the means available for solution are clearly laid out at the start. As Simon (1973) writes: 'an ISP [ill-structured problem] is usually defined as a problem whose structure lacks the definition in some respect. A problem is an ISP if it is not a WSP.'

Many classical studies of problem-solving are based around puzzles in which the goal is known. Thus the start point and the end point are set beforehand and all that is required is to find a path (in some appropriate space) from one to the other. Perhaps it is the over-emphasis on problems of this kind which has lead to the exaggerated importance of formal methods of solutions as a model of human problem-solving (Wason and Johnson-Laird, 1972). These classical studies seem to assume in effect that the diagnosis is already known and all that remains is to show that the evidence justifies it. They are therefore inappropriate as models of the general decision-making process. If a key feature of WSPs, as opposed to ISPs, is that in the former both the initial and goal states are known, then medical diagnosis generally seems to fall in the latter class, and some kind of transformation or formalisation is needed to change it into a WSP (Hand, 1985). However, in many ill-structured medical problems, once the initial states have been established and the medical diagnosis has been made then the treatment decision-making process generally becomes a well-structured problem.

Current expert systems, developed for the medical arena, have concentrated on the medical diagnosis problem. The developers therefore focussed almost entirely on the ISP of classifying the presenting symptoms at the beginning of the diagnosis. The selection of the treatment strategy is determined by the diagnosis and is well-established once the symptoms have been classified.

Conversely, in some well-structured medical diagnosis problems, i.e. ruptured knee ligaments, the decision-making process over treatment becomes ill-structured, since numerous treatments exist but their outcomes and effects have not yet been fully established. Therefore in this situation there is no formalised process of selecting a method to get from the known initial state to a desirable goal state.

These situations are not catered for by the current expert systems. For example in knee ligament injuries, it is relatively easy to identify which ligaments have been damaged and how bad the injuries are. The diagnosis is therefore relatively straightforward. However it is at this point that the consultant is faced with the difficult question of how the patient should be treated. Surgey or not surgery? Reconstruct with artificial material or the patient's own tissues or another animal/person's tissue? Place the leg in a cast or brace or nothing? Exercise continually or frequently or infrequently or not at all? When to start weight bearing? What post-operative exercise should the patient do until 'normal' life can be resumed? These are only a selection of the questions that consultants must consider. The answers to all of these are vital to the 'success' of the treatment and a wrong decision would seriously effect the overall result. Consequently, treatment selection is difficult and ill-structured.

Moreover, how in fact should 'success' be categorised ? 'Success' in a previous case does not guarantee 'success' everytime. Treatment for this type of problem is not like treating a disease, for 'success' cannot be achieved merely by identifying and eradicating the virus or germ. Each patient requires a particular amount of knee functionality for their 'normal' life and their perception of what is 'normal' will be different. Therefore a definition for 'success' must be determined in some form between the patient and the consultant before a treatment strategy can be selected.

The orthopaedic consultants at the LRI were constantly being asked to select treatment paths for patients presenting knee ligament injuries without them being able to predict the likely outcome. The lack of a formalised approach for treating particular knee injuries and the uncertainty over the outcome of the various treatment strategies, were in fact the factors that instigated the desire to computerise the collection of patient information. The belief was that once the data was computerised then detailed analyses could be undertaken to determine the influences of the various characteristics of the injury, the patient and the treatment itself.

The consultants realised that 'success' would have to be defined with the patient and this would vary from one individual to the next or even between patients in the same profession, i.e. the expectations are likely to be different between a top flight footballer who had 10 years ahead of him against another who was nearing the end of his career. However, by analysing the differences between patients, treatments and outcomes, insights will emerge into the relationships between certain characteristics and the likelihood of achieving the desired results.

These problems do not confine themselves to knee ligament injuries. In fact, most areas of medicine are unformalised. Consequently, the various consultants from the different fields need the necessary tools to analyse their domains, and thereby gain assistance in the decision-making process.

In these situations, current expert systems are of little use due to their structure and operation. They have been developed for a different problem area and are therefore not designed for such situations. However lessons can be learnt from reviewing expert system development. For example, the identification of the most productive knowledge acquisition techniques, the importance of end-user involvement throughout the design process and the crucial influence on the overall project of the user interface. But a vital question which previous expert system development cannot answer is what are the actual processes involved when humans are trying to decide upon a strategy to be adopted? The initial decision-support systems attempted to model the decision-making process without fully understanding it. What, however, is required is an identification of the weaknesses in human decision-making and whether appropriate computerised tools can be built to assist in these areas.

5.5 Decision-making

After defining the initial states, the next stage in decision-making is to identify the options available and project and evaluate the possible consequences of any of the choices taken (Evans, 1989). Decision-makers must therefore establish a set of objectives (or goals), whose attainment depends upon the decision taken. In some instances there will only be a single objective, though in many real-life situations there are a number of objectives and conflicts may occur in achieving acceptable levels of each. In these situations, decision-makers must search for a range of possible options from which a set of alternative courses of action (or strategies) can be determined. This search process is often difficult and may require decision-makers to contemplate various scenarios and strategy paths. A measure of the value or payoff (utility) of each possible outcome in terms of a decision-maker's objectives must then be established (Moore and Thomas, 1976).

The objectives, strategies, events and payoffs are the basic structural elements present in a decision problem. After structuring the problem the next stage is the analysis and the determination of the best strategy according to the desired goal of the decision-maker. As Collste (1992) states, the ideal judgement, 'presupposes impartiality and total knowledge of the situation and the consequences of the different alternatives. The action that leads to the satisfaction of as many preferences as possible is the one that should be chosen'. However, uncertain environments exist, for example in business or medicine, making the analysis of the decision situations highly complex. If a decision-maker knew with certainty the various outcomes, the problem would be readily solved since all the decision-maker would have to do would be to choose that strategy for the given outcome which would maximise the gains in terms of the stated goals.

Duncker (1926) believes that most solutions are achieved through 'resonance'. 'Resonance' is the largely automatic application of previous experience to the present situation - by means of cognitive-perceptual responses. These cognitive-perceptual responses are set off through reactions to 'signals' from the immediate environment in which the problem is set.

This proposal supports the theory of 'habit-family hierarchies' which have also been accredited as having an important role in problem-solving behaviour. 'When placed in a problem situation the subject exhibits certain behaviour directed towards a specific goal. These will be the most likely responses in view of the regularity with which they had been reinforced in the past. Once a specific response-sequence fails it is dropped and another emerges. Which particular response emerges depends upon its position in the hierarchy' (Aitkenhead and Slack, 1990).

A further term, 'recollection', is used to refer to the active process of setting up prospective retrieval cues, evaluating the outcome, and systematically working towards a representation of a past experience that is acceptable. 'Specific encoding operations performed on what is perceived determine what is stored, and what is stored determines what retrieval cues are effective in providing access to what is stored' (Baddeley, 1990).

These few theories alone demonstrate that all thinking depends on learning. The learning unit in the human must be able to form hypotheses and it must receive feedback to evaluate the hypotheses and revise them if necessary.

'Learning by discovery' is meant to include not only scientific research, but also many smaller-scale events in which someone formulates a hypothesis, gathers data to test it, and uses the results to adjust their 'theory'. It involves defining new concepts, formulating new heuristics, and even adjusting or changing one's representation of knowledge. Learning by discovery is much slower than other forms of learning, such as being told something, but Lenat and Feigenbaum (1991) believe it is the chief method that extends the boundary of an individual's knowledge. The point being that while having a theory is essential, it is equally important to examine data and be driven by exceptions and anomalies to revise, criticise, and if necessary reject the original theory.

In an inductive reasoning task one has to try to discover a general rule by inspecting samples of evidence. Such tasks are especially interesting in that they constitute an analogy of research in the natural science, where natural laws must be discovered by controlled experiments. The best way to achieve this is by forming and testing hypotheses about a general rule (Evans, 1983).

For specialists to examine a hypothesis or theory they must therefore gather relevant information using a standard method, i.e. pre-defined questionnaires. They can then query the data once it is collected and thereby test their theories by examining the outcomes. Case histories can thus provide the data and outcomes of previous decisions. Consequently, they can be used to test a hypothesis or provide further evidence to support a particular theory.

Many decisions taken by a physician are not 'certain'. In other words, there are very often cases where no single choice is possible, but the experience of physicians, their knowledge of previous, similar cases, and their ability to balance the evidence are of paramount importance in choosing the correct line of behaviour (Lesmo et al, 1989).

Medical students or young physicians, however, do not have such an extensive knowledge base. Pauker et al (1976) observed that students or house officers, apparently to counter this problem, often approached the process in a highly structured, methodical fashion. Furthermore, they also noted that experienced physicians performing outside their area of expertise used a far more structured approach than was their usual custom. Consequently, if a domain expert devises the questionnaires, clinicans who are less experiecenced in the domain can collect all the required information by following the appropriate paths through the questionnaires. Hence, less is left to chance, to the expertise of a physician in the domain or to human frailty arising out of, for example, tiredness or loss of concentration. Once all the relevant data has been gathered, any complex decision over the selection of the treatment strategy can be the responsibility of the domain expert. In this manner, the expert can utilise his knowledge and experience in the most vital part of the decision-making process.

Providing experts with tools to analyse case-histories and test theories quickly and easily is one aspect of this research. The belief is that if experts can use statistics to investigate trends and examine hypotheses then a clearer understanding of the specialist domain will evolve. Furthermore, conclusions drawn from other studies and teams, working independently, can be confirmed or questioned by carrying out the same analyses on collected 'home' case histories.

In addition, the data can be used by specialists to assist them in making a decision over the treatment of a new patient. The outcomes of previous cases with similar characteristics can be explored. Furthermore, by predicting the effect of the differences between a new patients' characteristics and previous cases, alternative treatment outcomes can be reviewed. From these reviews certain characteristics will demonstrate their importance in the overall result obtained whilst others will reveal that they have less impact. This will lead to more information being uncovered concerning the effects of an injury and subsequent treatment, and the influence and importance of the various patient characteristics.

Previously decision support systems attempted to computerise the actual decision process of the experts but made very little, if any, use of the available information in the case histories. However, human problem-solving proved to be far more complex and imprecise than was first anticipated and endeavours to model even the most restricted problem space required vast amounts of time and effort.

Decision-making is not a standard, straightforward process. Individuals have different values and views of the world. These have been established over time from the experiences and influences encountered by people during their lives. The following section gives a brief overview of some of the effects of such individual weightings and biases on decision-making processes.

5.6 Weightings and biases

Memory plays a major part in human decision-making. The results of past decisions and what has been learnt from them is stored in memory. Particular experiences in memory are likely to be found at the structures that were used to process them. The use of such structures is for prediction and explanation of future events based upon prior experience. Hence a person's previous knowledge of the task environment will exert a considerable influence during problem-solving with prior beliefs and expectations determining which pieces of information are considered by a problem solver (Evans, 1983).

Consequently, in problem-solving situations, the knowledge that guides the solution process is knowledge which a problem solver associates directly or indirectly with the problem at hand. For example, in the diagnostic process, a doctor will take into account the likelihood of each disease occurring in a particular patient. This initial belief will be effected by the environment, i.e. the likelihood of a patient having beri beri is different in the USA than in Tanzania. Hence, in the USA, doctors would not diagnose beri beri unless the symptoms were very evident, even though it is theoretically possible that a patient is suffering from this disease (Moore and Thomas, 1976).

The reason for this is that individuals perceive similarities and differences only because they associate characteristics at certain values with given events and situations (Garg-Janardan and Salvendy, 1988). According to Kelly (1955), the anticipated outcome depends on an individual's interpretation of past similar events. Problems are solved in this manner by drawing on the knowledge learnt and stored from solving similar problems in the past. Due to the role of previous experience and personal traits, the knowledge structures formed and stored differ from individual to individual. Hence, this leads to the same problem being solved in different ways by different individuals. Also, within the same individual, knowledge structures are continually modified and extended by incorporating the experience and knowledge gained from solving new problems. This leads to the consensus and knowledge acquisition problems outlined in Chapter 2.

Previous experience may indicate the presence of certain attributes in current alternatives which were also present in those past alternatives which led to pleasant outcomes. This association may lead humans to look for these attributes or rate them higher in future decision problems, irrespective of their relevance to the current problem. For example, people often predict by selecting the outcome (for example, an occupation) that is most representative of the input (for example, the appearance of a person). The confidence they have in their prediction depends primarily upon the degree of match between the selected outcome and the input, with little or no regard for the factors that limit predictive accuracy (Tversky and Kahnemann, 1990).

A great many sources of bias which hinder decision-making have been identified by cognitive psychologists (Fieschi, 1990). Two are the 'confirmation bias' and the 'belief bias'. Both of these biases reflect a tendency for people to try to maintain their existing beliefs. The difference is that while confirmation bias refers to a tendency to seek evidence which confirms one's theories (and to avoid evidence which refutes them), belief bias constitutes a biased assessment of evidence presented (Evans, 1987ab).

In addition, Fieschi (1990) lists a number of other biases which occur regularly:

- the 'inconsistency' bias which induces different advice on identical cases,
- the heuristics used to reduce the mental effort, such as the habit of choosing an alternative because it has been satisfactory in the past,
- the 'justifiability' bias in which a rule is applied if a person can find a reason to justify it even if it is not appropriate,
- the environment in which decisions are taken, which plays an important part and is susceptible to decision bias,
- the influence, on a decision, exerted by a majority expert group which effects the judgement of the minority, or stress or the manner in which a choice is requested.

Furthermore, there are situations in which people assess the frequency of an object or the probability of an event, by the ease with which instances or occurrences can be brought to mind. This is called the 'availability' heuristic. As Cohen (1987) describes, 'we over-estimate the probability of publicised events, such as lotteries; and students, for example, under-estimate the probability of dying of heart disease, since fewer instances come to mind. Availability introduces uncertainty about the accuracy of our assessments of probability'.

Therefore, events which are more accessible in the cognitive organisation of memory due to being more recent, more 'vivid' (salient), similar to prior beliefs and expectations and so on, may be available disproportionate to their true frequency (Evans, 1989; Schank, 1990). For example, the impact of seeing a house burning on the subjective probability of such accidents is probably greater than the impact of reading about a fire in the local paper (this would be due to the salience bias). Alternatively, Evans (1983) suggests, that subjects' retrieval may accurately reflect their experience, but that experience is subject to biased samples of evidence.

Furthermore, base rates have been widely shown to be neglected altogether by subjects making intuitive probability judgments in a large number of experiments, although there are some conditions in which they should be taken into account (Pollard and Evans, 1983). It has been suggested that this kind of faulty statistical reasoning can lead physicians to make irrational decisions, for example, to recommend unjustified investigative surgery.

It must be recognised, therefore, that the knowledge which experts bring to bear upon problems include beliefs which may shape their investigations and perceptions of evidence, thus introducing biases into their judgments. This means that the 'reasons' that the subjects give for their choices are highly misleading if taken to be 'strategy reports' (Evans, 1983). The biases reviewed here do not include motivation effects such as wishful thinking or the distortion of judgments by payoffs and penalties. Indeed, there are reports where severe errors of judgment occurred despite the fact that subjects were encouraged to be accurate and were rewarded for correct answers (Tversky and Kahneman, 1971; Kahnemann and Tversky, 1972).

Consequently, biases and weightings cause major problems for any system attempting to mimic the human expert. If, during the knowledge acquisition process, an expert is required to allocate probabilities to certain events or symptoms, there is a high likelihood that those values will be inaccurate. In certain systems these subjective values determine how the system behaves and the responses that it gives. The operators are therefore unlikely to have full confidence, if any, in such a system. Any answers based on incorrect information are suspect, whether these answers are generated by humans or by computers.

Moreover, as it has been explained previously, the biases and weightings held by a person are particular to that individual since they have evolved from that person's everyday experiences. Thus, if a system relies on the subjective values assigned by an individual, its acceptability by other consultants, and its transferability to other locations, will be reduced drastically. Consultants in each field would have to have a system built specifically for them. As experts' values change so too must the system. Consequently the development process would become a continual update-re-evaluate cycle, with little chance of being used for productive work.

The method this research advocates to attempt to eliminate the biases and weightings in human decision-making, is to use statistics to investigate patterns within the collected data. Through statistical analyses evidence can be gained to substantiate subjective views and beliefs or, in fact, reveal relationships previously unknown.

The literature suggests consultants tend to formulate a hypothesis early and then collect data to determine if the hypothesis is correct. Using the same technique they can devise questionnaires prior to an investigation and thereby ensure the collection of standard data. Furthermore, the specification of questionnaires would also assist more junior doctors or consultants working outside their specialist fields. As the literature showed, these physicians prefer to use a far more structured route through the decision-making process than the experts.

The use of standard procedures enables a computer to be utilised to carry out reviews and statistical analyses, question results and investigate links between data items and flaws in the diagnostic process. These analyses are far more likely to be free from the biases and errors of human analyses. They will however still be reliant upon the correct collection of the initial data, the selection of a valid analysis for the data items and the final interpretation of the presented results. These factors are, and will remain, the responsibility of the human consultant.

If the statistical analyses are properly conducted, insights can be gained into why one procedure succeeded whilst another failed and also why one procedure succeeded one time and failed the next. More importantly, the conclusions drawn can be justified and supported with evidence gained in a scientific manner rather than proposed as subjective views or 'gut feelings'. These conclusions can therefore be verified or questioned by other teams working independently. In addition, focussed discussions and debates can be initiated and the beginning of a more formalised, established specialist field can emerge.

Consequently, that aspect in which a computerised tool will assist the decision-maker needs to be outlined.

5.7 Identification of areas of the decision-making process in which assistance could be given

In Chapter 2, the limitations of humans as information processors was described. These restrictions prevent decision-makers from undertaking complete and comprehensive analyses of their information. However, if a set of computerised tools was available which could assist a decision-maker in carrying out analyses, this problem would be overcome. A system which could provide such facilities would enable a user 'to devote more effort to problem-solving by automating some of the tasks previously handled by the user', (Todd and Benbasat, 1992).

Conceptual models of decision-making have suggested that effort is a key factor in determining strategy selection (Todd and Benbasat, 1992). Hence, if everything is equal, a decision-maker will always try to minimise effort. Therefore, the assumption is that if a computer system could save decision-makers effort, they would re-invest the effort in problem-solving, thus producing better decisions.

In addition to this assumption there is of course a further issue, that of the acceptability of the computer system. If users perceive that a computer system will save them effort in their work, they are more likely to use the system since they will select the decision pathway requiring the least amount effort.

Consequently, the design of a decision aid should focus on these two issues:

- what aspects of the decision-making process can be automated to save a decision-maker effort,
- if automated, is the amount of effort required to use the system, i.e. in terms of interacting with the system and in understanding and processing the information generated by the system, substantially less than would be expended by an unaided decision-maker. If these factors are adequately appreciated, it is highly likely that the system will be used and will result in higher quality decisions.

Three effects of decision-aids on the decision-making process discovered by Todd and Benbasat (1992) were:

- a decision-aid appeared to replace some of the information processing activities of a decision-maker,
- decision-makers seem to switch their attention more towards problem planning and prioritising rather than concentrating solely on the problem at hand,
- a decision-aid appeared to make decision processes of decision makers more stream-lined and consistent. In this manner, possible alternatives were not 'forgotten', as appeared to be the case with unaided decision-makers. Hence work was not having to be repeated and alternatives were not being re-examined after they had already been shown to be unsatisfactory.

All of these traits increase the efficiency of the problem-solving process and the quality of the decision reached.

The conclusion drawn by Todd and Benbasat (1992) was that an appropriate decision-aid could reduce the information processing load for decision-makers and that decision-makers would then re-direct their efforts towards activities not supported by the decision-aid. Parsaye et al (1989) agree, adding that these systems can also be used to 'discover relationships that users would not have expected. Since today's oceans of data are abundant with these relationships, these tools will dramatically increase our ability to distil knowledge from databases.'

There is of course the assumption that the data on which a decision is based, either human or computer-based, is in fact correct. As Parsaye and Chignell (1993) explains, 'knowledge is power, knowledge grows out of information interpretation and information interpretation is founded on data. Bad data leads to erroneous knowledge.'

Good case histories are therefore crucial. If statistical techniques are to be used to analyse the information, the rules regarding the collection of data for statistical analysis must be strictly followed. For example, the use of standardised questionnaires, appropriate allocation of patients to the different groups in the trial and the similarity of the environment and conditions when patient data is collected. The use of standardised forms has already proven to be beneficial, for example, by reducing the amount of memory required to collect all the data, i.e. questions and procedures are not 'forgotten', by stream-lining and forcing consistency on the data collection process, and by assisting clinicians who have little experience within the domain.

If the information has been collected correctly, the ability to review and analysis it quickly and easily will greatly assist any investigator. Selecting by hand the correct data items from a mass of case histories is boring, mundane and time-consuming. As is carrying out statistical analyses by hand, by calculator or by entering data related to a specific analysis into a computer package. Paper case histories can be misplaced, lost or damaged and errors are likely to occur during the selection, transcription and calculation phases of the analysis. Consequently, a computer system which could provide the facilities to store case histories and permit the review and analysis of all the data held, would be of great benefit to an investigator, especially if the interaction techniques require little effort, no prior experience or learning and are quick to use.

Such a computer system would assist decision-makers during the analysis of their data and in testing hypotheses and theories. It would allow them to investigate possible patterns and relationships in and between data, thereby enabling them to understand and identify the importance of specific data items. Ultimately, these investigations will enable a more formalised field to emerge, where agreed standardised procedures are used and where guidelines have evolved against which other treatments, techniques, tests, etc. can be compared. New procedures, including treatments, would therefore have to explain the additional benefit each would bring in relation to these established guidelines. Such a process would also, hopefully, reduce the current ad hoc method of adopting techniques which have never been fully evaluated.

5.8 The relevance of statistics

Two common approaches to data analysis are: exploratory and confirmatory analysis. Exploratory analysis is the process of discovering patterns and relationships in data whereas confirmatory analysis is the validation process of showing that such relationships do in fact hold true (Parsaye and Chignell, 1993; Nabney and Grasl, 1991). As previously explained, humans tend to be very good at noticing possible patterns of relationships in or between cases. They are, however, poor at carrying out mental validation of their theories. Confirmatory statistics enable them to devise a hypothesis (stating that a mean of one set of numbers is significantly greater than the mean of another set of numbers) and then to carry out the statistical test to see if the hypothesis holds true. Using such techniques, decision-makers can explore the data in their field and begin to make their decision-making more explicit.

Previously, in medical fields, physicians attempted to define their decisions by devising algorithms. The belief was that structured decision pathways could be created to assist physicians, who were experiencing major problems with information overload. Kunz (1984) refered to the significant improvements which occurred when physicians used algorithms. Moreover these improvements disappeared when the physicians stopped using the algorithms. This seemed to indicate that the improvements were due to using the algorithms themselves rather than there being a purely educational effect.

One reason for the improvement could be that the algorithms increased physicians' adherence to detailed procedures, leading to the adoption of a very structured approach. In this manner, the limited information processing capabilities of the physicians were not being exceeded and the chance of 'forgetting' or overlooking an alternative was greatly reduced. This theory of the benefit of having a structured approach is supported by the conclusions drawn by de Dombal (1984) and de Dombal et al (1972) after he noticed the improvements in physicians using standardised, set forms for diagnosing acute abdominal pain. Therefore carefully designed algorithms, it seems, can improve the quality of decisions and thus patient care (Kunz, 1984; Szolovits and Pauker, 1978)

Two further advantages of algorithms are that:

- they can be paper-based and are therefore cheap and easily transported,
- they have been used by assistants, nurses as well as physicians, with similar beneficial results.

Consequently, in situations where trained physicians are scarce, i.e. in developing countries, or where much travelling is required to remote populations, algorithms can be of great assistance for straight-forward, well-defined, simple problems. However, for complex cases, algorithms possessess no means to undertake statistical analyses or to incorporate probabilities.

Statistics enable the data of a patient to be compared with a statistical model. If the fit is sufficiently close, a decision-maker can infer, with some confidence, that the patient has the properties of the population to which the patient most closely corresponds. This knowledge can then be used to assist in diagnosis or in treatment selection.

De Dombal et al (1972) developed possibly the best known computerised example of using statistical probabilities. In this system for acute abdominal pain, the data was entered from standardised sheets, summarising clinical and laboratory findings, became the attributes which were subjected to Bayesian analysis. The conditional probabilities of each of the seven possible diagnoses had previously been compiled from a large group of patients. The assumption was that each patient had one of these diseases and the most likely one was selected on the basis of the recorded observations. The program achieved an accuracy equal to that of human experts. Moreover, the system's performance improved as the number of patients increased.

Leaper et al (1972) carried out an investigation in which the performance of de Dombal's system, based on conditional probabilities calculated from 600 patients, was compared with the accuracy achieved by another system based on estimates of conditional probabilities from experts. The system with the experts' estimates performed less well than the unaided clinician. However, the system based on the real-life data was significantly more effective than the unaided clinician (hence also supplying supportive evidence for the findings of de Dombal). These results provided further justification for the belief that individuals possess inaccurate weightings of evidence and poor recall of specific situations or events.

Evans (1989) claims that medical diagnosis is, in fact, a statistical decision process of classic Bayesian characteristics. He states that the posterior odds in favour of the condition considered (i.e. odds after viewing the evidence) are a function of both the prior odds (or base rate) and the likelihood ratio of the evidence. The base rate probability of a patient suffering from a condition is relative to the knowledge about the patient. For example, the prior probability of a heart attack in someone who is male, middle-aged, overweight, and a heavy smoker is considerably higher than that of the average member of the population. The likelihood ratio of the evidence refers to the possibility that the evidence could arise due to the suspected cause relative to the chance that it could arise from any other cause.

However the Bayesian technique can not be applied effectively to all medical domains. It has a number of limitations:

- the assumption of conditional independence of symptoms. If this does not apply, there can be significant errors in the results obtained (White, 1985),
- the assumption of mutual exclusiveness and exhaustiveness of disease categories. Often there are concurrent and overlapping disease categories (Shortliffe et al, 1984,
- the requirement that:
  - the conditional probabilities are stable over time, i.e. the effect of an antibiotic remains the same,
  - the techniques used to gather data are the same,
  - the effect of the variations in probabilities over geographic location and population are limited.

Bayes' theorem is only one technique from a much wider field of decision analysis. In general terms, decision analysis is a method by which values associated with choices, as well as probabilities, are considered when attempting to analyse a decision process. For example, the best solution may not be adopted if the price is too high, both in terms of its actual financial cost and also in terms of the impact it has on a patient, i.e. if a test procedure is very painful, a definitive diagnosis might not be sought, similarly a patient may not wish to undergo an operation for a minor complaint.

Hence, a rational decision includes cost-benefit analysis such as considering financial costs, patient discomfort, possible compilations and patient preferences. An 'expected value' or utility is calculated for each pathway from the probabilities and the costs and the 'value' of the outcome. The ideal solution pathway is the one which maximises the expected value.

At one time this technique of using expected utility maximisation had much support for medical decision-making (Lindley, 1985; Schwartz and Griffin, 1986). However, from Lindely's (1985) description of the process, the difficulties are evident, 'first the uncertainties present in the situation must be quantified in terms of values called probabilities. Second, the various consequences of the courses of action must be similarly described in terms of utilities. Third, the decision must be taken which is expected - on the basis of the calculated probabilities - to give the greatest utility'. In other words, all the decision solutions must be clearly pre-defined and a comprehensive set of probabilities and cost/benefit parameters established for the whole situation.

This can be very difficult to accomplish (Fox et al 1990), especially as the a priori probabilities possessed by experts have been proven to be so unreliable. Therefore automated decision analysis requires well-defined, limited problem spaces where statistical analysis of real data can establish the required probabilities. The true costs and the value of the outcome may however be harder to determine when dealing with medical fields. Here it has been suggested that subjective probabilities should be used and that these should then be analysed over time to produce more established probabilities (Savage, 1972; Gage, 1993). However, when the various probabilities have been defined and Lindley's criteria have been satisfied, the few decision aids that have been built have performed well and their performance has degraded smoothly if the quality or availability of data happened to be reduced (Fox and Krause, 1992).

Consequently, the need for a large amount of correctly collected standardised data is very important to such mechanisms. Even though patient data are generally variable for a particular disease at a global level, Szolovits and Pauker (1978) believe that when the disease is subspecified into a particular complaint, severity, or age, such effects become less variable, with many becoming nearly invariant. This is also likely to apply to treatment outcomes and strategies.

There has been the realisation of the importance of collecting large amounts of data however little has progessed from the early reports. For example, Shortliffe et al (1984) described a national project 'to obtain enough data so that groups of retrieval patients will be sizeable, thereby controlling some observer variability and making the system's recommendations more statistically defensible'. Kunz (1984) also agreed with the importance of gathering such data through organised methods, explaining that, 'uncertain decisions can be analysed with Bayes' theorem using probabilities derived from the results of treating the local hospital population'.

Therefore statistical analyses could be very useful aids in medical decision-making, in researching into prognostic significance of tests and in evaluating the effectiveness of treatments. In this manner patient discomfort, time expended and financial expense are all reduced, thereby improving patient care and the efficiency of health care provision.

Probability theory is once more starting a revival as a means of representing uncertainty in decision-making (Parsons, 1994). The belief is that when complex problems need to be addressed, e.g. which treatment to select, probabilistic models are needed. However, as Szolovits and Pauker (1978) warn, 'the essential key to their correct use is that they must be applied in a limited problem domain where their assumptions can be accepted with confidence.' Gage (1992b) also believes that mathematics can represent uncertainty in clinical practice, 'in the same fashion that quantum mechanics precisely captures the uncertainties of our knowledge of elementary particles'. However, it must be remembered that mathematical models alone can not be expected to be sufficient to make final decisions. Rather they can assist in removing some of the uncertainty inherent in human decision-making, thus leading to better, more efficient and effective decisions being made.

5.9 IDDA analysis section

Generally an investigator will wish to:

1) view the data stored in the databases, e.g. review the records of a particular patient,
2) compare the data using a variety of statistical techniques,

for example, find the number of patients who have undergone a CBF (carbon fibre) operation,

determines the mean Anterior Drawer Laxity of each of the operations,

determine if there is a significant improvement in knee laxity after a CBF operation.

Initially two methods were proposed for the analysis phase of the IDDA end-system. Both are outlined briefly in the following sections.

5.9.1 Original proposal

In the original proposal, a definition for 'success' and 'failure' must first be established, both in subjective and objective terms. Also, from the statistical analysis of past data, the factors influencing the outcomes would have to be determined. These factors would then be ranked, with their exact or range values, with respect to a new patient's details, and matched to the initial assessments of previous cases, thereby producing a list of outcomes.

For example:

Let 'success' be defined as a final knee laxity <=0.1mm and the patient suffering 'No Pain', and, 'failure' as being classified as final knee laxity >0.5mm.

If statistical analysis revealed that the outcome of a treatment was influenced by:

1) the age of the patient,
2) the sporting level,
3) the sex of the patient,
4) the initial activity score achieved,

then these influencing factors would be linked to the new patient's details, with any allowable ranges assigned, and a comparison of the past cases carried out. For example, if the new patient was 24 years old, had a sporting level of 4, was female, and had an initial activity score of 10, all previous cases where the patient was between 22 and 28 years old, with a sporting level of 3 or 4, female, and with an initial activity score of between 8 and 14 would be selected and their outcomes compared with the 'success' and 'failure' classification.

Hence the 'success' rate and 'failure' rate of each treatment with respect to the above definitions would be revealed e.g.

Treatment Ranking	No. of 'successes'	No. of 'failures'
1)-------------	Exact Fit;Best Fit	Exact Fit;Best Fit
2)-------------etc.		

'Best Fit' is the number of patients with the influential characteristics similar to the new patient in the allowable, user specified range whereas 'Exact Fit' is the number of patients with exactly the same influential characteristics as the new patient. It should also be noted that the 'failure' classification would indicate the number of occurrences that each particular treatment had been deemed to have failed with respect to the definition, and not all the times that the treatment was tried but not classified as 'successful'.

This technique, however, was believed to be too narrow to prove to be truly beneficial to the doctors using it, e.g. the results would just present 'success' and 'failure' and not how each treatment performed in each of the different tests carried out during the assessment stages. Also, the words 'success' and 'failure' were not appropriate as doctors are likely to feel uncomfortable using a system that seemed to be rating their performance.

A further problem was the danger that ranking the treatments would imply that one treatment was better than another when only a selected number of criteria had been analysed. For example, a patient might wish to have a treatment that requires less convalescence, though the result is likely to be that the knee would be looser than if a different treatment was used which required a longer period of recovery. Also the effect on the patient's social and working activities have to be considered before a final decision over treatment is made.

Consequently, it was believed that this approach would provide:

- incomplete advice, since cost-benefit analysis had not been incorporated and yet the results presented gave the impression of a final ranking, i.e. that a final decision order was being listed,
- too narrow a focus, as only a comparison of a selection of new patient details against previous cases could be undertaken rather than allowing an investigator to select the data items and the technique to be used in an analysis,
- inappropriate assistance for an ill-defined, unformalised domain since definitions for 'success' and 'failure' had to be established at the outset as had the patient characteristics which were deemed to be influential in the outcome of treatment strategies,
- the possibility of small numbers of patients in each category, especially at the outset of a study or if a large number of influential characteristics were defined for the domain or stringent 'success' and 'failure' definitions were made. This is likely to result in inaccurate and incomplete information being presented to decision-makers,
- facilities that are likely to be of very limited value to domains outside the knee ligament field,
- a system that is highly likely to be rejected by its intended users, Not only for the reasons stated above, but because the results obtained reveal very little information that would be of benefit to

a decision-maker. Facilities to investigate the possible treatment paths and patient characteristics in more detail are required, especially in an unformalised, ill-defined domain, i.e. provide specialists with the ability to analyse and investigate their own fields in the manner they specify.

Consequently, a different approach was considered.

5.9.2 Revised Proposal

In the revised proposal, doctors determine and select the statistical analyses that they wish to undertake. A global condition can be specified, if needed, to narrow the selection of records for an analysis. For example, by specifying that only those records with patients who have: age = 20-30, and gender = female should be used. These global definitions will be of the influential characteristics within the field. (However the determination of which characteristics are in fact influential will have already been established from previous analyses on past cases, i.e. the same analysis tools can be used to carry out a specific analysis to determine whether the gender of a patient influences or effects treatment outcome.) The criteria for the particular analysis to be undertaken are then specified. For example, the different treatments against knee laxity after the operation. The analysis is carried out and the results are presented in an appropriate format.

This method :

- allows more criteria to be analysed,
- allows different analyses to be undertaken,
- enables consultants to control and determine the analysis, thereby permitting them to decide which hypothesis or theory they wish to test and the method to be used,
- requires no subjective definitions for 'success' or 'failure',
- does not appear to review consultants' performance, thus is likely to appear less threatening,
- utilises known and accepted techniques for data analysis, hence the methods used are easily understood and the results can be easily checked and verified. Thus the dissemination of results and conclusions via journal publications is relatively straight-forward,
- enables more detailed data analysis to be undertaken, thereby assisting in both the formalisation of the field through standardising techniques and procedures as well as furthering the medical knowledge of the specialist field through revealing relationships within and between data,
- permits simple comparisons to be undertaken at the outset of a study when only a small number of patients may exist, i.e. finding the average age of the patients in each of the different treatment groups, etc,
- allows the IDDA system to be applied to any field which carries out statistical investigations,
- enables the identification of tests and procedures which provide little, or no, extra information for the decision-making process, thereby reducing patient discomfort, financial costs and consultant's time by justifying the removal of such procedures,
- enables similar studies to those conducted at other institutions to be undertaken and the original findings to be confirmed or discrepancies to be uncovered,
- provides assistance in a task that investigators will require, i.e. any study that is undertaken will at some time be analysed statistically. Therefore it will save an investigator time and effort, if

statistical facilities are built into the database system. Moreover, if the specification of the statistical conditions is intuitive and the presentation of the results is in appropriate format, users will find the analysis section easy and quick to use and understand. Hence the acceptability of the IDDA end-system will be enhanced.

Therefore it was decided that this method would provide the users with better facilities with which to investigate their fields, especially unformalised, ill-defined domains. However, in the same manner that the design of the data entry section of the IDDA system was important, so too is the design of the Analysis section.

5.9.3 Design decisions

The factors that were reviewed in Chapter 3 were also influential in the design of the analysis section of the IDDA end-system. For example, the proposed users, the task to be undertaken and the environment in which it is to operate, etc. Hence the current commonly used practice of expecting an investigator to 'dump' the details from the database in an appropriate format for a statistical package to read and then to have to switch from the database system to the statistical package to load in the data, specify the analysis, identify the data required and attempt to decipher the results and information presented, was considered to be inappropriate for the scenario expected for the IDDA system.

Therefore the operation of the analysis section was designed to be as close as possible to the operation of the rest of the IDDA end-system, i.e. the database section. The collection and manipulation of the data for the specified analysis would be hidden from users as would the initiation of the statistical package and analysis test.

Thomas (1987) undertook studies of users carrying out queries and made the following recommendations:

- have users select rather than produce the queries whenever feasible,
- be sure users know how to solve problems, is familiar with the database, knows how to use the I/O and knows how to get help when needed,
- have potential users participate as fully as possible in the design and implementation of the system,
- have users write the query directly in a format that reflects the intrinsic data organisation.

From these suggestions and from the literature reviewed in Chapter 3, it was decided that the interfaces for the analysis section would be based exclusively on the menu selection and form filling styles. This would therefore enable users to quickly and easily define and select both the statistical conditions required and the statistical test to be undertaken. All the interfaces follow the same model no matter which statistical test is selected. Consequently, for intermittent, naive users the ease of use, learnability and acceptability will be increased by this standardisation of the interface.

In addition, as the analysis section will be used by domain experts, who have also designed their system, users should be able to solve problems and should be familiar with the information stored and the domain itself. As explained in Chapter 4, the interfaces have been designed to be as similar as possible to the other sections of the system and therefore the interaction techniques are standardised as well, i.e. based solely on the keyboard with two screens available - one for help text and the other for data entry. With regards to forming a query, the next section will describe this process in more detail. However, it has been designed to be as intuitive and straight-forward as possible. The other facilities that are available at the Analysis stage are also explained in the following section.

5.9.4 Facilities available during the analysis section

In the statistical analysis stage, the availability of the second screen is invaluable. Similar to its role in the design and development of an IDDA, it displays help when requested by users without obscuring the main entry screen, which is displayed on the main monitor. Therefore the user can view both the help screen and the entry screen simultaneously, thus enabling cross referencing to the help and the examples whilst completing the entry screens.

Furthermore, the second screen plays an important role when users are specifying the data sets for the analysis. Domain experts will know the assessment questions which acquire the data they wish to analyse. By specifying the appropriate question numbers from the assessment questionnaires, conditions and rules can be built to enable the required analysis to be carried out, i.e. if the various treatments are to be compared then a user selects the questionnaire that contains the question which asks for the treatment to be specified and then selects the question number associated with the question asking for the treatment to be defined.

However, rather than users requiring to refer to paper lists, or remember the question numbers, it is more efficient if users could view the assessment questionnaires before making a selection. By using the second screen, on which the questions can be summarised, questions can be viewed without interfering with the data entry screen. Once users select a question number, the question is displayed in full on the main monitor, thereby enabling users to check that it is the correct selection before defining the condition that this question must satisfy for the analysis. Hence, any errors in question number selection can be quickly detected and corrected.

The second monitor is only ever used as a reference screen and all actions and responses are always carried out on the main monitor. This should therefore reduce any confusion that might arise from switching screens.

To define and manipulate data users can construct conditional rules that represent the problem to be analysed. As the literature showed (see Chapter 2), experts can make associations between pieces of information and can propose theories to account for patterns in data. Therefore, it is reasonable to assume that they can determine the conditions they wish to use to query the data. To enable the required conditions to be specified, there is the ability to produce compound conditions, i.e. only consider female patients who have had a carbon fibre operation should be considered. Consequently, conditions can be ANDed and/or ORed together to construct the necessary data sets.

The literature indicates that problems seem to occur in the definition of the statistical conditions. Wason and Johnson-Laird (1972) found that subjects could more easily describe conjunctive concepts (involving AND) than disjunctive ones (involving OR) and that concepts which involved both conjunctive and disjunctive relations were the hardest to describe. Essen et al (1991) agree, 'the AND operator was used as a connective in only one third of the electronic searches; the OR and NOT operators were never used'.

It is believed that one reason for the difficulty stems from an incompatibility between natural English usage of 'and' and 'or' and their use in database retrieval. Ogden and Kaplan (1986) investigated this problem in detail, showing that the English word 'or' is most frequently used to indicate union, but that 'and' is often used ambiguously to indicate both union and intersection. Thus the statement 'Show the students in grades 10 and 11' implies the union operation (the set of students in grade 10 plus the set of students in grade 11), despite the use of the word 'and'. Since people often attempt to translate the English-language statement of a problem into the query language phrase-by-phrase, the different meanings of the connectives are not always taken into account.

Moreover, Wason and Johnson-Laird (1972) found that subjects seemed to prefer positive descriptions of concepts rather than negative ones, even though the negative description could be more efficient. There has been considerable literature concerned with people's ability to comprehend positive and negative statements, in which the latter are almost universally observed to cause extra difficulty (Wason, 1980; Evans, 1982). Therefore it was decided that a NOT operator would not be included as an option during the definition of a condition.

However, in an attempt to help clarify compound conditions, there is the ability to specify parentheses. This facilitiy is also required to ensure that the correct, and expected, comparison is made. For although AND and OR have execution priorities, the contents of a set of parentheses over rules these priorities and ensures that the conditions inside are executed first. The literature reports that the ability to use parentheses to separate out the conditions greatly assists subjects during a condition definition. For example, the analysis by Essens et al (1991) showed that the effect of parentheses reduced the amount of processing needed in mixed category conditions, since processing of the second operator can sometimes be omitted. The parentheses seemed to help the subjects mentally segregate the processing of the two operators, resulting in a large reduction of errors.

Consequently, it is anticipated that this ability to include parentheses will help the users whenever they are required to define compound conditions. The IDDA does check that an even number of parentheses has been specified and also suggests where the parentheses should be placed, though the actual inclusion is determined by the user.

Two facilities available at every entry stage of the analysis section are the ability for the user to request help and the ability to return to the main analysis menu immediately, i.e. escape back to the main menu after cancelling the analysis.

The help that is displayed will be relevant to the analysis being undertaken and the user entry required at that particular instance. In addition to this help within a particular analysis, there is also explanations of each of the analyses with examples available at the main analysis menu, i.e. before selecting an analysis users can obtain descriptions of the various statistical tests and review the worked examples to satisfy themselves which will be the most appropriate test for the analysis they wish to carry out. As explained above, all of the help is displayed on the second monitor to ensure that users can still read the data entry screen whilst referencing the help text.

Certain analyses require that interval values be defined for one or both of the conditions. At these entry points users have the facility to re-specify an earlier interval value if they detect that a mistake has been made.

A further facility at this stage enables users to display a single question condition on the second monitor. This gives users the ability to reference the full question and answer lists during the interval specification.

Therefore, users have a number of facilities available during the analysis section. These facilities were selected to assist users in areas that the literature and previous experience have highlighted as causing problems. It will only be through sustained use of this section of the IDDA system that evidence will emerge as to the usefulness of these current techniques or whether in fact further facilities are required to help users as they attempt to analyse their data.

Consequently, the design of the statistical interface is very important. Users who are not statisticians, and who may well be infrequent users, must be able to select the desired analysis and specify the required data sets quickly and easily. The importance of statistics as a technique to examine collected data and test hypotheses makes its inclusion within medical systems essential. Statistics enables results to be presented in a more formalised and acceptable manner. This ability is crucial to fields such as medical decision-making where so much of the actual processing is reliant upon subjective judgements. Hence, it is intended that this research project will provide tools to enable the decision makers to review and investgate their field and to help analyse and justify their hypotheses and theories. The interface for the statistical tools, therefore, has been designed to be easy to understand, quick to operate and provide help to the user whenever it may be needed.

5.9.5 The available analyses

There are a variety of statistical analyses available in the IDDA system. These are:

the Sign Test,	the T-Test,	the Wilcoxon,
the Mann-Whitney,	the Chi-squared,	the ANOVA,
Regression,	Correlation,	F-test.

In addition, the user can specify histograms to be drawn by defining the condition for the X-axis and the condition for the Y-axis, or by just defining the condition for the X-axis, causing the Y-axis to display the number of observations in the various categories. For example,

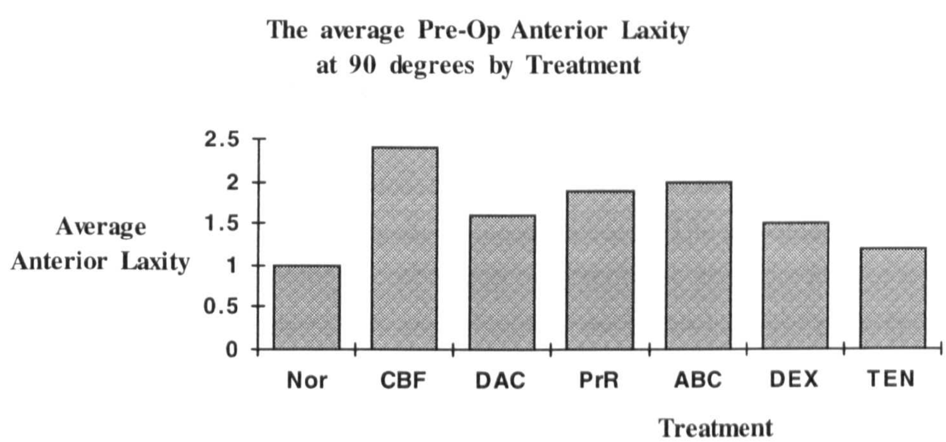


Figure 5.3: An example of a histogram with both axes defined by a user

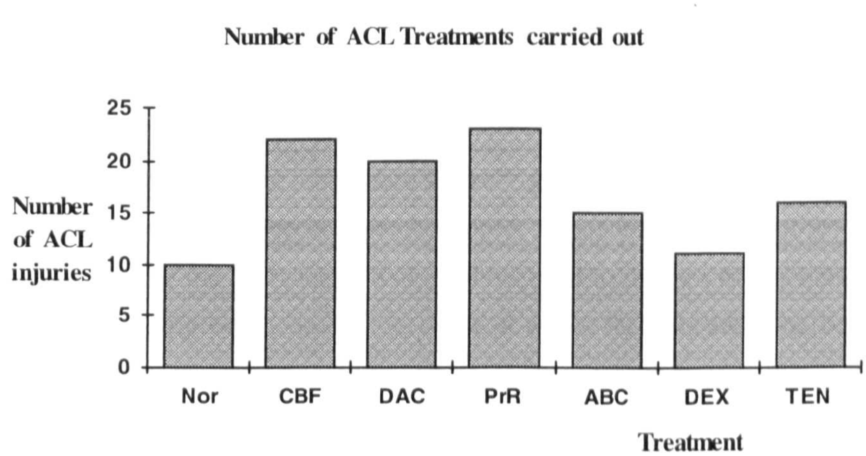


Figure 5.4: An example of a histogram with just the X-axis defined by a user

If both axes are specified by a user, a display of a table holding the mean, standard deviation and number of observations in each category can be requested. In either case, i.e. after defining just the X-axis, or both axes, a user can request that the various category case-study numbers are displayed. Consequently, a hand check can be carried out, if required.

Once a user has identified the type of analysis to be undertaken and the conditions under which the data is to be collected, the system gathers the specified data and carrys out the selected analysis. The result is then presented to users. All of this takes time and does not require any further user interaction. Consequently, status messages are used to display to users the progress of the data sorting and testing. The actual amount of time taken is dependent upon the number of records in the database, the number of conditions to be reviewed and the type of test to be carried out.

This delay, however, is a fraction of the amount of time required for the same analysis to be undertaken by hand. In addition, the computerised approach requires a fraction of the effort on the part of a user, is more accurate and more thorough. Consequently, not only can the results be relied upon to demonstrate the findings of the selected analysis correctly, but many more analyses can be undertaken in the same period of time than was previously required to carry out a single analysis by hand. Using an IDDA will therefore save users time and effort, with all the subsequent benefits.

It is from the results obtained from these analyses that indications of associations and links between data can be highlighted and further investigations can be initiated. The presentation of the results therefore also becomes very important. As the output from statistics packages tend to be very brief, the results from an analysis can be rather confusing. Therefore the IDDA captures the output from the statistics package and extends it before displaying these result to users in a format more applicable for non-statisticians. The IDDA system however does not attempt to fully interpret or draw conclusions from the results. It displays the outcomes from an analysis in an easy to read form but leaves the final interpretation to the users themselves. No questionnaire or statistical analysis can fully capture all of the factors that will influence a final decision. Therefore, a computer system can only hope to offer some assistance to decision-makers to enable them to make more informed and hence better and more justifiable decisions than they could possibly make on their own.

It must also be remembered that while computers are very good at storing and retrieving facts and comparing data, a computerised system is dependent upon the accuracy of the information entered, the search techniques selected by the users and the algorithms used to derive conclusions. Its "number-crunching" capabilities are impressive and this is a task at which humans are poor, but currently the computer lacks the ability to interpret and discriminate, tasks at which humans excel (Coates and Vlaeminke, 1987). Users must therefore be aware of the scope and the limitations of computers, without which users cannot fully understand or appreciate the information being presented nor use the system effectively. As Anbar (1987) pointed out, 'using a thesaurus of synonyms improves the richness of your style and allows you to express your thoughts more precisely, by reminding you of words you may have forgotten; but, if you do not know the language, you may end up in real trouble and would have been better off not using the thesaurus in the first place'.

Moreover as the literature showed in Chapter 3, no matter how efficient a computer system is if it usurps the user, or is deemed by the users to do so, the system will not be used. However, the IDDA is a tool and has been designed by the end-users for their specific task. It stores the information users have decided is important and it undertakes the analyses users have selected and defined. The final interpretation of the results from any analysis and their applicability to the situation under review are the responsibility of users. Therefore, users can investigate theories or trends without the tiresome and mundane tasks of sorting the records, gathering the relevant data, and then actually applying the required tests, before finally arriving at the result. These are left to the IDDA system whilst users can propose further investigations or new theories, draw conclusions from previous analyses, apply these conclusions in a practical way, consider the consequences of new techniques and procedures in their domains, or, determine the applicability of the old methods. Consequently, the IDDA does not impinge on users' intelligence or primary skills but assists in the tasks to which a computer is best suited. By placing these additional methods and techniques at the disposal of end-users, an IDDA end-system attempts to assist them in their tasks.

To summarise, therefore, an IDDA end-system merely enables users to collect information in an orderly fashion, to review and analyse the details quickly and easily, thereby enabling them to extend their knowledge of their field in a easy, quick and more thorough fashion than is currently available to them. When users operate the system, it is under their control as are the investigations undertaken, the conclusions drawn and the effects of applying any subsequent actions. Hence, the users are the ultimate decision-makers and, as such, determine the actual influence that an IDDA system will exert upon their work and the whole environment in which it is to operate.

The following chapter discusses various evaluation techniques before describing the summary of the results obtained during the evaulation of the outcomes of this research.

Chapter 6

The Evaluation Process

6.1 Introduction

At best, the evaluation of medical expert systems, particularly those intended for decision support in the clinical domain, has been an ad hoc process. Few developers have placed sufficient emphasis in this stage of the development cycle. In fact, 'at present there are very few systems which have been rigorously validated by well codified and guaranteed methods. Generally the performance of existing systems is evaluated in special circumstances which introduce biases that distort estimation of their capabilities in actual cases, as well as the expertise level of the knowledge base' (Fieschi, 1990).

Wyatt (1987) discovered from a cursory review of 14 papers on artificial intelligence in medicine, that only 3 adequately described the system's clinical role, 7 had a scientific test set analysis, and one was subjected to a field trial. Lundsgaarde (1987) found similar results, 'approximately 90% of all automated medical expert systems have not been independently evaluated for performance in controlled or real-time clinical environments'. The reason, it is claimed, is that such evaluations are difficult and time-consuming, especially those involving clinical trials, (Wyatt and Spiegelhalter, 1990; Reggia and Tuhrim, 1985).

In medicine, however, evaluation is especially important if the safety of patients is to be entrusted to such systems. Doctors mistrust computers and have, in the past, shown considerable resistance to their introduction. Therefore doctors need to be convinced that automated systems can offer them anything other than merely a mechanism which, indirectly, questions their integrity and restricts their clinical freedom. For example, the process by which a system reaches a decision or result must be comprehensible to users who must agree with the methods used to reach those results. However, as already stressed in Chapter 5, experts often reach decisions by different routes and often arrive at different conclusions. Consequently, if it is hard for experts to agree amongst themselves, it is even harder for a computer system, constructed from one expert's knowledge base, to convince a different expert that the methods and pathways it uses are correct. Hence difficulties arise as Fieschi (1990) notes, 'when expert and system agree, it may merely correspond to an error of judgement. When the system, disagrees with the expert, how often is the expert in error? Finally, it should be pointed out that it is difficult to assess system performance in rare disease cases, the limited number of such cases introduces evaluation bias into the system'.

Responsibility is a key issue. Not only will the performance of a medical system be important in determining whether a doctor was right to use one in a case of alleged negligence, but, if an American judgement of 1981 (Wyatt, 1987) is upheld, doctors may be found negligent if they do not use a computer system that has been shown to improve performance. Therefore, there is growing pressure to derive agreed criteria against which a system can be assessed to the satisfaction of both users and developers.

Lundsgaarde (1987) states that it is misleading to claim that existing techniques for evaluating systems are few and primitive. He believes that it is much more accurate to say that the majority of developers of expert systems often ignore the many human, contextual, and cultural factors that determine whether a new system will be accepted by the end-users. This ignorance, indifference or lack of perspective, is likely to be a major factor influencing why only a handful of expert systems have made the successful transition from laboratory to the clinical implementation. The difference in numbers between research developments and real-life installations is still evident as Sorgaard (1991) notes, 'there is a disparity between the multitude of apparently successful expert system prototypes and the scarcity of expert systems in real everyday use'.

Computer systems should evolve as direct responses to the needs of developers and users. The needs of these two groups may be very different and may sometimes be in direct conflict. Therefore, users and designers are highly likely to apply different criteria when they assess a system since, with different goals, they will perceive the system and its facilities differently (Booth, 1991). As Flagle (1982) discovered, 'to our delight the program worked perfectly, at which point one of the visitors said 'That's an interesting system, have you evaluated it ?' In shocked fury my colleague responded 'What do you mean, evaluate - it's working, isn't it ?' To my colleague, avoidance of fiasco was an adequate criterion of effectiveness at that moment, while to the site visitor, value could be found only in some demonstrably improved service or health outcome'.

Therefore, for an expert system to be successful, it must also be acceptable to the users and hence its 'usability' and benefits are very important aspects for study and evaluation. However, many usability evaluations are made against ad hoc criteria rather than against criteria that were established prior to the evaluation and used as a basis for design (Chapinis and Budurka, 1990). Shackel (1986) proposes that an operational definition of usability should be used to set the criteria. This definition incorporates: effectiveness, learnability, flexibility and user attitude. In contrast, Clegg et al (1988) suggested six issues that should be considered when assessing usability: ease of learning, being in control, degree of effort, system speed, getting information in and out and errors and their correction.

However, usability is only one factor; there is the issue of building systems so that they fit the working practices of the unit which they are intended to support. The assistance must be supplied at the appropriate moment, in the appropriate manner and with a suitable tool. Hence, understanding the intended task, the working environment as well as the users, is very important. This is true not only in defining the design requirements (as stressed in Chapter 3) but also in establishing evaluation criteria. If there is early specification of the latter then a prototyping approach to system development can be adopted, as there are known criteria against which the initial prototypes can be compared, i.e. a design, evaluate, re-design cycle can be followed.

The effects of the system on the user's activities and their environment must be assessed. This includes looking, 'not only at its effects on direct users but also at its effects on indirect users. These are people whose day to day work is affected by the introduction of a system even though they do not directly interact with the system' (Berry and Hart, 1990).

There are, in fact, many ways computer systems can contribute to the progress of medical science and to the improvement of health care, for example: improving data collection, establishing explicit management strategies for typical situations, giving comfort to the physician, allowing the browsing of medical knowledge, permitting the retrieval of representative cases, and enabling the analysis of past cases to provide evidence for supporting or refuting theories, thereby assisting in the making of better and more informed decisions in the future.

However, for computer systems to fit into pre-existing organisations, there will be some unavoidable modification of the tasks and roles that previously existed. The system could lead to permanent changes in users' attitudes and in the overall organisation of their work. Rossi-Mori et al (1990) describe the possible long-term effects as being:

1) on the positive side: providing a sort of permanent education and enabling deeper insights into the field and better data collection methods to be used,
2) on the negative side: developing a form of excessive trust and 'passive agreement' with the normally correct behaviour of the system, that eventually may lead to boredom or lack of supervision. For example, if a system generates reports on the interpretation of a laboratory test, which are correct for 99% of the cases:
    a) will doctors maintain their attention whilst checking the output of such a system?
    b) will they pass it to non-medical staff?
    c) will they be so lazy as to avoid the correction of less significant inaccuracies?

However as Berry and Hart (1990) point out, 'given that so few expert systems have been in regular use for a long period of time, there is little objective data about the long term effects'. Consequently, evaluations need to review far more than purely technical issues of speed and accuracy, as the end-system will impact upon the users', their environment and the tasks that they undertake. These issues must therefore be considered before a full picture of a system's effect and influence can be truly determined. The next sections describe the various aspects of the evaluation process.

6.2 When to evaluate

The evaluation process covers both general and specific issues. However, the precise timing of evaluations within the system design, development and implementation cycle is not specified. Hewett (1986) terms evaluation of an already built system 'summative' and evaluation which occurs as part of the design process, 'formative' evaluation. He argues for iterative 'formative' evaluation, which he regards as the important driving force behind the process of successful design. The results from system evaluation, during any stage of design, determine the direction of the design and the changes required during redesign. This type of cycle can be construed as 'design-evaluate-redesign' (Johnson and Johnson, 1988; Berry, 1994). An evaluation process, therefore, must be structured to create and maintain a clear focus on the goals of the project. Berry and Hart (1990) agree that, 'in very recent years there has been a trend towards a more thorough integration between design and evaluation, with an appreciation that evaluation should occur throughout the development process'.

Evaluations which occur after a system has been built can however also provide useful information. They have two main functions at this stage - one is to compare the proposed system with those already built, the other is to produce recommendations for the construction of a new system which would build on the advantages, but cut down on the disadvantages, of the old system.

With medical systems, evaluations are also essential to demonstrate user acceptability and in certain cases safety and ethicacy. Berry and Broadbent (1987) provided evidence that without a clear understanding of user needs and requirements, system builders will fail to provide the crucial capabilities needed. This would result in limited system utility and the high likelihood of the system being rejected by the intended users. Therefore there are various points in the implementation process where evaluations should be undertaken to enable the necessary user feedback.

The basic theme is that if a system cannot be evaluated, it is senseless to take it into use. However, any evaluation becomes absurd when performed without a formulated goal or when performed on a system which has no stated functional objectives (Brender and McNair, 1989).

6.3 What should be evaluated?

Before evaluation can start, an adequate description of the system and its role is required. Some basic questions should therefore be considered, for example as Nykanen (1989) suggests:

- in what context is the system planned to be used, by whom and for what purpose?
- what kind or type of system is needed?
- what specific requirements do users have with respect to terminology, reliability and security?

If these types of question are answered then a problem specification for the system will be defined. Wyatt (1987) believes that a protocol could then be developed describing:

- how the correct or 'gold standard' solution to the problem is to be derived,
- the potential users of the system and their current performance at the task,
- the amount, quality and types of data available for input,
- the output required,
- the level of explanations needed.

In addition, the actual 'context' should be described: this includes the area where the system will be used (GP surgery, A/E dept. etc.), the implications of the system's advice, and finally, a judgement should be made regarding the significance of errors. Once the problem specification and the system's role have been specified, there exists criteria against which the system's performance and impact can be compared during the evaluation process.

There is however the question of the user's work process. Will work be delayed by the system and can the system be integrated into the daily routine? How the system operates and it's influence on working practice are critical issues to the practical applicability of the system. As Berry and Hart (1990) point out, 'a useful system is one that helps users to achieve their goals. A system may be easy to learn, effective for the tasks it addresses and provoke good user attitude ratings. However if the functions of the system do not match the users' goals in their everyday work environments, then the system will not be used'.

Yet as Rossi-Mori and Ricci (1988) discovered, most of the evaluation of medical expert systems refer to just 'the final result' and use human experts in the field as a 'gold standard'. Murray (1990) agrees that most evaluations only seem to address the scientific question of whether the diagnosis or prognosis is accurate, whilst stopping short of answering such questions as whether clinical care is improved by using the system.

Therefore if speed and accuracy are the only measures used in evaluation, the system is highly likely to be rejected by the users. To address this problem, Rossi-Mori and Ricci (1988) propose four levels of possible evaluation:

1) raw efficiency of the system in itself: the 'technical' performance of the system with respect to defined requirements and to a specific problem, apart from a user's general work context;
2) effectiveness in a user's environment: the performance of the system in the environment of a specific class of users;
3) long term effects on a user's behaviour: the acceptance or rejection of the system in the routine work and the features of the assimilation of the system's advice in the decisions;
4) effectiveness to the medical problem: the impact on the specific health problem that the system is addressing, the acceptability of the system to a wide range of users and to the more general issue of the global impact of computer-based systems in the health field.

6.4 Where should a system be evaluated?

Often systems are not evaluated in a user environment before they are released for general use. Therefore, unfortunately an evaluation environment can often be very different to an intended user environment. For example, they differ with respect to the problem mix, and how familiar users are with the program. Consequently, evaluations outside a user environment can only be rough approximations of the results expected when the program is placed in a users' workplace. Moreover, it has been the case that many expert systems developed in laboratories have never even reached the stage of formal evaluation (Lundsgaarde, 1987; Buchanan and Shortliffe, 1984). Instead, they have been caught in continual iterations of the refinement stage, i.e. looping round for further alterations to be made. This is possibly another reason why so many expert systems have never managed to move into widespread, 'real-life' clinical settings.

It is clearly insufficient to invest resources in a system because it works to the satisfaction of the developer or system expert. It is essential to establish how well a system works in relation to the functions it may complement, improve, or replace. Therefore from the very beginning issues of a users' environment must be taken into account. This is particularly relevant in the field of medicine. Rossi-Mori et al (1990) believe that many projects failed because 'the human side of medicine' was affected by using the system: 'we are sure that every attempt to interfere with the patient-physician relationship will produce a failure, even in the most authoritative expert system.'

They add that the interaction of users with the system should be compatible with a viable reorganisation of users' activities. For as MacLean et al (1985) state, relying only on time measures as a metric for the evaluation of the interface and the software may involve hidden costs in terms of user acceptance. For example, fatigue or frustration may quickly result from forced use of an interface which places too heavy demands on prospective users. Consequently the user interface is of vital importance to the acceptability of the system, as stressed previously in Chapter 4.

6.5 Transportation of systems

Hilden and Habbema (1990) suggest that properties, which they classify as 'non-frozen', should also be considered when evaluating a system. For example, 'on the technical side we have portability, maintainability and the property of mergeability; and on the medical side, transferability to new medical environments and updatability as medical science progresses and new generations of diagnostic and therapeutic procedures are born' (Hilden and Habbema, 1990).

As more computer systems are developed, a necessary facility is the ability to transfer a system from one location to another. This requirement effects how the system is designed, since changes in operational procedures must be accomplished with ease and versatility. If a system is to be transferred to another place, i.e. a different institution, it must be designed to allow for any necessary alterations to be carried out without major time delays.

Problems in the movement of systems from one area to another have been reported, especially within the medical field. Some of these have been highlighted by Bjerragaard et al (1976), Horrocks et al (1976) and Adams et al (1986), who studied the moving of the statistically based acute abdominal pain system constructed by de Dombal et al (1972) from Leeds to Copenhagen, Airedale, and a further eight centres. The findings were that the overall accuracy was lowered since the disease presentation varied in the new areas and also because other diseases were classified as "acute abdominal pain". This meant that certain diseases had not been accounted for in the original system and the prior probabilities needed to be altered. Hence systems that perform well in one situation may not be able to transfer that achievement directly into another environment.

Other problems occur when transferring medical expert systems as there are no standard procedures for treating diseases, and the diagnostic procedures that patients go through, differ from one hospital to another. Hence, a system that has been accepted by one group may be totally inappropriate to another unit. For example, a rigid system which insists that test 1 should be carried out before test 2 will be rejected on the grounds that it would require changes in normal hospital practice. Consequently the design of systems should be as flexible as possible to enable alterations to any of its parts to be made easily and quickly without effecting the overall reliability, accuracy or robustness of the end-system.

Furthermore, variation exists in the terminology used by experts within the same field to describe and classify observations and results, and also in the phrasing of general questions. These differences have serious implications on the design and flexibility of the user interface of the system. The interface has been shown to be a crucial factor in the acceptability of the system to end-users (Hudson and Cohen, 1985; Berry, 1987). This problem could, thus, be a limiting factor in the transferability of the final application, thereby restricting the benefits that could be gained from the widespread use of such a system.

Currently, systems have tended to have been built for only one particular location, hence not all the problems and difficulties that exist have come to light nor, consequently, have the solutions. During the design and evaluation of a system, however, transportation is still an issue that should be considered since it has a significant influence on how systems should be designed, developed and built.

These sections demonstrate that the criteria used for evaluation exceed those of the purely technical properties of speed and accuracy. Specifications are required to describe the system, its role, its impact on its users and the environment in which it is to operate. A system's usability will be evaluated for its compliance with these specifications, and thus human factors become an important and influential aspect of the main stream of development and evaluation. The next question therefore relates to how these criteria can be effectively measured.

6.6 How can a system be evaluated?

Evaluation and validation are necessary tasks to perform before taking any system into routine use. The key principles of evaluation, according to Gould and Lewis (1985) are early interactive involvement of users, empirical measurement and iterative design. Evaluation may focus on a number of aspects of design, for example usability, learnability, acceptability and functionality. Watts (1987) identified a number of key features of software quality as being: correctness, maintainability, usability, reliability, portability, security, efficiency, reusability and interoperability.

Whilst evaluating is measuring quality characteristics, validating is checking the quality and comparing quality measures with a frame of reference. However, as yet, there are no strict procedures for either (Valluy et al, 1989).

Miller and Sittig (1990) describe measures of merit as typically being in subjective-objective pairs. For example, perceived versus actual gain in diagnostic performance, perceived versus actual time savings, reassurance versus objectively justified reassurance. This is useful to remember during the planning of an evaluation and the studying of selected criteria, as the interpretation of results will depend upon the evaluator's view as to what precisely is being assessed.

In addition, developers have discovered that users are not always as interested in the question 'Does it work?' as 'How does it work ?' - the pragmatic and the explanatory motives for evaluation. Consequently, the real purpose for performing an evaluation should be clearly stated from the outset, as this dictates what is to be measured (Wyatt, 1987).

Some criteria, such as the speed with which decisions are made, the reliability of the system in terms of down-time, and the amount and quality of data that a system uses to make its decision, can be measured relatively easily (Fox et al, 1980). Others such as the satisfaction of the users with the system and the transparency and ease of maintainability are currently difficult to evaluate as there is no metric or 'gold standard' with which to make comparisons.

Wyatt and Spiegelhalter (1990) proposed a three component evaluation method:

Stage 1: consists solely of a definition phase. This should identify: the task to be undertaken, types of user, and users' physical and social environment. It may not be possible to achieve a complete definition of user requirements without first building an early prototype system and requesting users to comment on how this could be improved.

Stage 2: at this stage, formal laboratory testing can begin. There are two major groups whose interests can be served by this testing phase: the users and the experts who sanction its use.

Stage 3: as such systems only affect patients indirectly, by influencing decision-makers, field trials will determine if they have value to patients and/or to the decision-makers.

Generally evaluations focus on testing a single decision aid, first against the intuitive standards of the designers, later against unaided clinical performance. However, as Hilden and Habbema (1990) state 'comparing two or more CDSSs [clinical decision support systems] calls for some additional ingenuity in the study design'. This is possible, for though there has been an analogy made with testing of new drugs, there is an important difference, as Habbema et al (1981) point out: 'you can run two competing systems simultaneously but you cannot administer two competing drugs to the same patient.' Consequently, decision aids are and remain non-invasive diagnostic devices - regardless of whether they may suggest or trigger invasive procedures.

6.7 Users and the user interface

An influential, and in fact vital, component of a system is the user interface. It is tempting to identify the human-computer interface (HCI), as that which the user sees, touches and hears. However, it is insufficient only to investigate the entities with which the user interacts. For as Edmonds (1990) points out, 'to be worthwhile we must also consider just what the user can do with the system and how the user can do those things: which methods they can employ'. Once attention is drawn to users' methods the full environment in which they work must be reviewed because, almost always, their activity involves a mixture of methods, some of which will not be computer-based.

HCI evaluation should be seen as a contribution to the design process that ensures the quality of the resulting systems in relation to their use in practice, i.e. HCI evaluation can best be seen as design for operation. Its influence should start at the beginning of the design process and continue through to delivery. Feedback from users and from the application situation is constantly used to inform design decisions and to adjust and correct requirement statements. Edmonds (1990) believes that this feedback is in fact more important than the provable reliability of any particular experiment or evaluation conducted.

One of the most convenient methods of evaluation would be to match a design against a set of guidelines that indicated generic requirements for human-computer interfaces. Unfortunately, the state of the art in applied psychology is not at the stage that makes such guidelines easy to provide. Most of the guidelines that are provided relate only to the presentation layer of the interface, covering such issues as text format, the use of upper and lower case and appropriate colour combinations. For users, however, a very important aspect of user interfaces and human-machine interaction is whether or not they are able to operate systems with confidence and correctness. Work practice and environment both have a considerable influence upon the methods selected by users.

Furthermore, it must be established whether a system fits into the organisation, the working environment and the decision-making process with which it is to be integrated, 'the system must take into account the knowledge people bring with them to the task, in terms of both the actions and the objects they are performed on, and also in terms of the sequence of carrying out the various sub-tasks that make up the task' (Johnson and Johnson, 1988). In this sense, the evaluation of the dialogue control aspects of the system must be made in the particular application context in which the system will be employed. Here the concern is with operational requirements, i.e. a failure to consider these early on can lead to inadequate designs. Consequently, there are limits to the kind of systems which can be made, or to how good systems can be made, without involving end-users and other people affected by the system.

Evaluating a system from the users' perspective means measuring attitudes and 'user satisfaction', which are difficult to do formally, especially when the system being tested is merely a prototype. In this case, Wyatt and Spiegelhalter (1990) suggest that an informal study of users' attitudes and activities when exposed to the prototype should be undertaken as this should reveal important and useful feedback. However, they also warn that attention should be given to the correct timing of a trial, to avoid premature trials that might discredit the system and dishearten the developers and delay the implemention of an effective system.

Weighed against this, of course, is the risk of developing a system that is inappropriate for the tasks it is to carry out and/or one which the users will not accept or use. Therefore Sorgaard (1991) suggests, 'the idea is to start with a minimally structured system, for example, based on manipulation of free-text by an editor, and later on impose structure as patterns of use evolve. In other words, start with a simple and flexible, but useful system and see what happens'.

6.8 Summary of concerns affecting evaluations

Evaluation is an important part of the development life cycle of any computer-based system. In most applications it is generally the last stage to take place, however it should be an on-going and iterative process. It can be used as a means of reviewing and managing the progress of a project and also as a method of highlighting problem areas and concerns at an early stage. The ultimate evaluation should be carried out with the end-users.

It is hard, however, for a person who is not computer literate to review critically the performance of a system. To attempt to address this problem, de Dombal (1983) outlined, and expanded upon, the guidelines that Lusted (1976) suggested for non-experts to evaluate objectively an expert system. He stressed that the accuracy of results, though important, is not the sole consideration when reviewing a system. A crucial factor is the suitability of the whole application to the problem domain and thus extensive evaluation of all the aspects, and the effects of the system, must be reviewed thoroughly.

Only if systems are wanted, are usable in the working environment, and draw conclusions that seem reasonable to users, will they be accepted and used. This may seem obvious, but systems have been developed that failed because they were too cumbersome to operate, asked too many questions in an unintuitive order, took up more time than was available, or occasionally came up with answers that were clearly wrong, but for which they had no explanation (Wyatt and Spiegelhalter, 1990).

Lundsgaarde (1987) believes that one of the biggest barriers to successful development resides with the resilience of human beings (experts or non-experts) to learn and benefit from past experience. One possible reason for the failure of developers to take evaluation seriously might be 'perhaps for obscure reasons, there is less academic prestige in testing as opposed to building systems, and less prestige in developing evaluation tools as opposed to design tools' (Hilden and Habbema, 1990). However, without evaluation, systems will not be used and thus their benefits will be lost. The scale of possible benefits which can be achieved is shown in a study undertaken by Jydstrup and Gross (1966), who discovered that one fourth of the operating costs in three hospitals was directly related to information handling. With the present day exuberance over the collection of data and information, the economic cost of non-computerised information processing would now greatly exceed even this figure.

Unfortunately, the evaluation, which is required before a system can be widely used, is time-consuming, but it should still include a definite statement about the contribution of the system to medical practice. In addition, current evaluation techniques and practises need to be standardised in order to avoid the continued use of the current ad-hoc methods of assessment. Guidelines on testing and reviewing results would not only assist designers, but they would also aid end-users. The operators, who are quite often not computer experts, will then be able to participate in assessing the performance and effects of the computer system.

Moreover, these standard checks could be used by the legal profession as indictors of the efficiency, effectiveness, and accuracy of a final product. Legal issues are becoming a very important topic as medical systems and tools become more popular and their use more widespread. Consequently, careful evaluation of medical systems is a vital part of the development of working systems that can offer enormous benefits to patients and physicians alike.

6.9 Criteria for the evaluations of the tools and the IDDA system

It is believed that the major factor determining the success of the proposed methodology for the current project will be the acceptability of both the tools and the IDDA end-system to potential users. Therefore the tests and discussions undertaken during the design and the evaluation stages need to consider the issues of:

- the intended user group,
- the working environment in which the system is to operate,
- the tasks in which the systems could assist the user,
- the tasks which the systems would not undertake,
- the integration of systems' tasks and users' tasks to accomplish the overall goal of undertaking investigations within the specialist field to enable better, more informed decisions to be made and to enable the decision pathways to be understood.

Hence, the suitability and usability of the tools and the IDDA end-system and their likely effect on the users, the tasks, the working environment and the specialist field are, in this project, crucial to the overall acceptability of these systems.

Therefore, the questions of when, what and how to evaluate the tools, which construct IDDA systems, and IDDA systems themselves, had to be established. Following are the questions that were deemed to be critical for this project. These issues were considered during the iterative design and development stages e.g. from the outset. Brief summaries of the answers to these questions are included and cross-references to previous chapters are given. The other factors were evaluated on the completed tools or IDDA system and these have been described in more detail in Appendix C.

Certain long-term effects will not become evident until clinical trials utilising the IDDA system have been running for a number of years and investigations have been carried out on the results of these trials. However, the potential impact of such a system on health care and specialist fields was considered and discussed with potential users.

1 When to carry out evaluations, e.g. summative or formative?

As previously explained in Chapter 3 and Chapter 4, the project development cycle used a formative evaluation approach, i.e. a design-evaluate-re-design technique. This enabled not only the robustness and correctness of the code to be tested but also the feedback from appropriate users groups to be considered during subsequent development work. During later evaluations, however, the IDDA system was compared with existing systems to determine whether its design addressed users' needs more effectively than than did current systems.

These two stages in the evaluation of this project can be compared to the first two levels of the four level model proposed by Rossi-More and Ricci (1988) (described in section 6.3). Moreover, it is intended that the third and fourth level described in that model will be assessed during long-term evaluation of the tools and the IDDA system.

2 What is the suitability and usability of the tools and the IDDA system:

2.1 in what type of environment is the system to operate?

This was considered during the design of both the tools and the IDDA end-system (Chapter 3 and Chapter 4). As the trial environment was to be in a specialist medical unit, the common factors and difficulties experienced in such situations were reviewed, both in the literature and during discussions with the users of the LRI knee ligament system.

The findings were that the daytime environment is noisy and cramped with frequent interruptions for the user, who is also working under severe time constraints. During the evenings and at weekends, there are fewer interruptions, less pressure on time and the general environment is quieter and more relaxed.

Consequently during the daytime, the systems can not impose any additional workload on the specialist unit nor can they delay the unit's daily activities in any way. In some situations, the use of paper questionnaires during patient consultations may be required. Here the patient details are recorded on paper and later transcribed into the computer system. In other situations, the consultants may choose to enter the patient details directly into the system during the patient session. In this manner, the system is not perceived as being a threat to the patient-doctor relationship, as the users can determine when the system is utilised. The IDDA end-system must therefore be designed in an appropriate fashion to accommodate either scenario.

Currently, the analysis or review of the information collected in a trial is generally undertaken during the quieter periods of time, e.g. evenings and weekends. Even when the details are computerised, these investigations are still likely to take place at a time when the user is better able to concentrate with fewer interruptions. Therefore the design of these sections of the IDDA system does not need to address all the same environmental factors as the data entry sections. However, for interaction consistency and familiarity, many of the decisions taken for the data entry section will have an impact on the design of the analysis and review sections, e.g. minimising the amount of user interaction. Others will obviously have direct impact, e.g. cramped conditions with little 'free' space makes mouse operations difficult. Further discussion of this and other environmental considerations can be found in Chapter 3 and Chapter 4.

The initial evaluations of the design decisions to address these concerns were carried out during the prototyping iterations. At first these involved the system developer, then other internal researchers, before trialing with the potential user group. Final conclusions however can not be drawn until long-term evaluations have been undertaken in a number of different 'real-life' user environments.

2.2 can the systems be transported to other units or locations?:

2.2.1 what type of configuration is needed, is it readily available, and is it expensive?

It was decided at the outset of the project that both the tools and the IDDA end-system should be able to run on a hardware configuration which was readily available and thus relatively cheap to buy, maintain and upgrade. This decision was taken as the cost of the system was seen as one possible prohibiting factor in gaining user support at the start of a project; for example the users would not be interested if the overall cost was too high nor if there was a requirement for expensive specialist equipment that could not be used for other tasks. This concern was particularly appropriate for consultants working in medical units in the NHS.

Consequently, the hardware chosen was a PC, i.e. 386, 2Mb RAM, 80Mb hard disk, or better. A higher configuration would be advised if the PC was going to be used for other tasks that required Windows applications. Nevertheless to run the tools and the IDDA end-system, a relatively basic configuration would suffice. There is however the need for a dumb terminal to be attached via a serial link. The reasons for this second terminal have already been outlined in Chapter 4 and Chapter 5.

Similar concerns were addressed during the choice of software. dBASE IV for DOS was selected for the database package as it provides good functionality at low cost. It is also a package that is already popular and thus has a large customer base, including some hospital information service units. Hence supporting material, extra utilities and experienced programmers are readily available. Moreover dBASE permits the execution of other applications from a dBASE program and provides facilities which allow for easy import and export text of data files into and out of a database. Minitab was chosen for the statistical analysis sections. It provides all the analyses that are often utilised by medical consultants during trials. It is already commonly used in the medical domain for such investigations. More importantly from a developer's view point, it is small, e.g. it does not require much disc space for its programs nor does it occupy much memory during execution. It can also be run through a macro script and it easily reads in data from a text file and produces output to a text file. It is also cheap in comparison with other available statistical packages. Both packages are also readily available, relatively inexpensive and can run on a basic PC configuration.

2.2.2 can the configuration be used for other daily tasks?

As has been explained above, this question was considered during the selection of both the hardware and software. The hardware could certainly be used for a wide variety of tasks appropriate to a basic PC. The software could also be used for other tasks. In some hospital units, they are already using dBASE applications developed by computer programmers and MINITAB is commonly used for analysing data. Consequently during the choice of both hardware and software, such issues were considered as well as the applicability of the selections to the current project.

2.2.3 are major alterations required in the source code when it is transported to another unit or institution?

This requirement for the ability to transport applications from one institution to another, or for similar applications to be built in another institution, was one of the main reasons for this project. There was a need for computerised clinical trial databases with integrated statistical capabilities and this demand was obviously not restricted to one institution nor one specialist field. Hence, as a solution was required which could address different specialist medical domains, it was necessary to build tools which could construct the required end-systems rather than concentrate on producing one end-system that was applicable to one specialism and one institution and then monitor how it faired in other situations. It was also not appropriate to attempt to accommodate all the different specialist terminology and the different modes of operation within one system. As explained in the literature review in Chapter 5, the transportation of medical systems from one institution to another has caused problems in the past because of these difference in terminology and operational procedures. However, these issues have been addressed in this project by the decision to develop a suite of tools and to use these to build the required end-system.

Therefore the source code of the tools does not change as they are moved to other institutions. The new institution or unit provides the questionnaires for the trial that they wish to undertake. Hence, these questionnaires will be in the specialist terminology of the domain and follow that unit's operational practices. The tools use this information and other details that the users supply to construct an appropriate IDDA end-system. Consequently, the problems that have arisen in the past because medical fields are not generally standardised or formalised and thus have different working practices and terminology even within the same specialist domain, actually have little impact on the methodology proposed in this project.

Factors such as the usability of the tools and the applicability of the IDDA end-system must obviously be addressed after widespread, pre-longed testing.

2.2.4 what is the time delay before a working system is available in the new institution or unit?

With the proposed methodology, the time delay will be dependent upon the members of the specialist unit, e.g. the length of time it takes to devise the trial questionnaires and to operate the tools to build the required IDDA end-system. Long or questionnaires with numerous conditional questions will obviously take more time to devise and process than short, simple questionnaires. As questionnaires are commonly produced when undertaking controlled trials, the extra time for initiating a computerised trial only really involves the interaction time with the tools. The design of the tools did try to minimise the interaction needed whilst still ensuring that enough information was gathered to successfully build the IDDA system. However a clearer idea of the time required will materialise when a number of evaluations have been carried out after the finished tools have been used to build a few IDDA end-systems from the same questionnaires but with a variety of different users.

2.3 what tasks are being complimented, replaced or improved?

This question was considered during the design. Is it better to build a system which reviews a new patient's characteristics and then selects the 'best' treatment path? Or is it better to build a system that can assist a consultant by carrying out the essential but long, mundane, and simple tasks? If a system is to control the task and inform the user of the best course of action, difficulties arise in trying to obtain, understand and model appropriately the decision making processes of the current human experts (as previously explained in Chapter 2 and Chapter 5 as well as the user acceptability issues that were raised in Chapter 3).

Consequently, it would appear that a better solution would be to build a system which assists during investigations by removing mundane tasks from investigators, increasing the accuracy of results and allowing for more numerous and thorough analyses to be undertaken in a smaller period of time. In this manner, user acceptability of an end-system is enhanced whilst the difficulties of attempting to represent human decision-making processes are removed.

Therefore the IDDA system has been designed to assist consultants to explore their field by:
- storing appropriately the data in the field,
- enabling the review of the stored data,
- retrieving specified patient details,
- statistically analysing the data selected by the user,
- drawing simple bar charts to visually represent the results of user specified queries.

These facilities enable previous outcomes and decisions to be examined, thereby assisting in the production of better more, informed decisions being made in the future. It can also be used to investigate current tests, procedures and practise that are carried out in the field and to determine their relative merits. The result of such investigations could be a more standardised and formalised domain.

The task of the tools was to take the type of information which was commonly available in clinical investigations, i.e. questionnaires, and with minimal extra information from the domain user gather together enough details to construct an appropriate IDDA end-system. Therefore the tools were to replace the need to employ a computer programmer to build such a system, thus reducing the expense and increasing the convenience to, and overall control of, the specialist domain investigator, i.e. an IDDA end-system can now be constructed quickly and cheaply and when it is required.

The design of the tools and the IDDA system were therefore influenced by the tasks each were expected to accomplish. A review of the performance of each over a period of time and in a number of different domains, will be required to determine whether these objectives were in fact met.

2.4 *is the system appropriate for the tasks it is to undertake?*

2.4.1 *is the correct sequence of operations followed and correct terminology used for the task and working environment?*

As already stressed above and in Chapter 3 and Chapter 4, these two factors are particularly important for an end-system to gain user acceptability and thus be used. It has already been explained that building tools, which could utilise the specialist information from the domain, had been selected as the mechanism to address these issues. The questionnaires, devised by the domain expert, were to be the major source of information for constructing an IDDA system, for example, by providing the correct terminology, specifying the data to be collected and the actual patient tests to be carried out and finally by sequencing the tests and procedures in the correct fashion for the unit. Hence, the ability of users to understand, learn, and accept the system, is enhanced whilst minimising the reorganisation of the normal, daily activities of the unit in which the system is to operate. Determining whether such a mechanism has addressed these issues appropriately and whether the impact of the system is as expected, can only be uncovered after carrying out a number of different evaluations in different environments over a period of time.

2.4.2 *is the system reliable?*

Again, the reliability of the IDDA system will need to be reviewed over a relatively long duration, as actual investigations must be undertaken and enough patient data collected before the system can be fully evaluated in all of the tasks it can perform. Obviously initial testing can be undertaken with data from the LRI knee ligament system but this is only one specialist domain with a limited number of users.

The reliability of the tools can also be reviewed initially, i.e. can they build appropriate IDDA systems for a number of different scenarios? However, in a similar fashion to the IDDA system, their weaknesses can only really be revealed after widespread, long-term examinations of their operation in different domains with different user groups.

Consequently, although these initial evaluations have been undertaken, the assessment of both the tools and the IDDA system will continue when they are operating in real clinical settings and over a sustained period of time.

2.4.3 *is the amount, quality, type of data for input and output appropriate?*

There was a design decision at the outset to attempt to minimise the amount of data that users of the tools would need to enter, when building the IDDA system. Error checking has been incorporated to try to identify mistakes (as explained in Chapter 4), thereby attempting to ensure the quality of the user's entry. The type of data the tools require is dependent primarily on the questionnaires since the questions asked determine the answer characteristics.

With the IDDA end-system, the amount and type of data to be entered is also determined by the questionnaires. Similarly the quality of the entry is monitored by both type and range checking and by enabling the entry of 'no value', e.g. 'no value' can be used when no measurement or details have been recorded for the question. Obviously during analyses, any 'no values' are discounted and thus the results are not affected by omissions in the data collection process.

The output of the tools and the IDDA system has also been carefully considered since the presentation of information is crucial to a reader's understanding. The amount, type and quality of data output are all important and the design decisions for these and other interface issues are described in more detail in Chapter 3 and Chapter 4.

However, the evaluations of the tools and IDDA system, which include user feedback, demonstrate that the correct design decisions have been selected for both the input and output of information.

2.4.4 *what is the level of explanation that is needed?*

Currently, the amount and type of explanation and help that is available have been aimed at assisting infrequent, naive computer users who are familiar with the specialist domain but lack significant statistical and computing knowledge. The evaluations throughout the project indicate that they have in fact addressed this user group in an appropriate fashion.

2.4.5 *is the expected contribution from the users appropriate?*

This question queries whether domain experts can devise appropriate questionnaires to undertake an investigation in their fields. Chapter 2 has reviewed this issue and indicated that this should not be a problem.

With regard to the effort and contribution required by users to operate both the tools and IDDA systems, these factors have been assessed. As explained above and in Chapter 3 and Chapter 4, design decisions taken during development stages considered these issues and attempted to minimise them both during the building and running of an IDDA system.

2.4.6 *does the system utilise the knowledge that the user already possesses?*

Throughout the building and the operation of the IDDA system, there has been the intention to use to best advantage the knowledge that domain experts, domain assistants or secretarial staff already possess. The methods adopted to encourage this have been described in the various sections of this thesis. This ability was considered an important factor as it would influence the acceptability of the system to users, e.g. if users feel that the system is 'familiar' through the use of known terminology, questions and operations, they gain confidence, understanding and has less inhibitions in exploring the facilities the system offers. Whether too much has been expected of the users of the tools and the IDDA system is a consideration that must be reviewed and is included in the list above.

2.4.7 *has the system evolved from the users' needs, i.e. do the users want assistance in the tasks and do they want that help in the manner that the system provides?*

The IDDA system did evolve from the expressed needs of a group of orthopaedic consultants. It has been designed to address those tasks in which the consultants had indicated that they needed assistance. Only after a more widespread introduction, over a period of time, can an analysis be carried out as to whether the proposed technique appropriately addresses all of the tasks. It is possible to compare the current LRI knee ligament system with a knee ligament system built using the tools and to gain the opinions of the LRI users as to whether the new system would also be appropriate to their needs. Moreover, it is possible to compare a system developed by hand by another programmer for another trial with a system generated using the tools from the original questionnaires. Through these types of evaluations, initial feedback can be gathered of the views of users. Also indications of other tasks in which users would like assistance may emerge.

2.4.8 *was the design influenced by feedback from the users?*

A prototyping approach was adopted for this project to enable user feedback to influence subsequent design decisions. This whole process was considered to be very important during the development of both the tools and the model of the IDDA end-system and it is described in much more detail in Chapter 3 and Chapter 4.

2.4.9 *is the interface appropriate for the user group, including considerations of any cultural factors, tasks and external influences that may be involved?*

Similar to the issue above, the design of the interface for both systems was deemed to be a critical factor in the success of the project. Careful consideration was given to the interaction style used and the facilities offered to ensure that they were appropriate for the proposed user group, the tasks and the environment. Chapter 3 and Chapter 4 review these issues in more depth. The adoption of the prototyping approach did assist in obtaining user feedback regarding the initial interface designs. However, further evaluations will be needed with different user groups and in different environments to determine whether the interfaces are still appropriate in all the conceivable scenarios in which an IDDA system, and the tools, may operate.

All the above issues were therefore considered during the design and development stages of the tools and IDDA end-system. From the evaluations that were undertaken as the project evolved, the feedback gathered influenced subsequent design decisions. With respect to usability, the objectives were to build systems which users : could understand, would find easy to learn and use, would feel confident in using and would feel in control of the task and the computer system. Moreover, the system was to operate at an appropriate speed for users, i.e. neither too fast nor too slow, provide error reporting and error recovery, and require minimal amount of effort, both physical and mental.

A further factor considered was the effect that the systems may have, for example:

1 are the working practices and daily activities affected?

2 what are the implications of the output from the system, i.e. will it lead to excessive trust or passive agreement, what is the likely significance of any errors?

3 is there any impact on health care or the specialist medical field, for example:

3.1 by undertaking tasks that users previously found difficult or were unable to do?

3.2 by saving time and/or money?

3.3 by enabling investigation to be undertaken on the decision-making process?

3.4 by permitting easy browsing of medical data and information?

3.5 by allowing rapid retrieval of past cases?

3.6 by assisting in the establishment of explicit management strategies within medical fields?

3.7 by undertaking user specified analyses of past cases to asssit in gathering evidence to question or support theories?

3.8 by improving data collection?

3.9 by helping to evolve a more formalised, standardised specialist medical field?

3.10 by improving the acceptability and credibility of computer systems within medicine?

The majority of these issues require not only the initial evaluations undertaken during the development phase but also additional testing over a longer period of time, in different working environments, supporting different specialisms and with different users, to establish whether the decisions taken and the solutions proposed were appropriate.

Nevertheless, for all evaluations there needs to be a decision regarding how the evaluations are to be carried out and recorded as well as selecting the users to take part in the tests. As explained above, the decision was taken to use an iterative prototyping life-cycle thereby utilising user feedback and discussions from the intermediate evaluations to focus and direct subsequent development work. The methods used to record information were:

- through this iterative development cycle which included user feedback and informal discussions,
- through observing users during their interaction with the system,
- through audio-recording of users during their interaction with the system,
- through audio-recording of users post-hoc comments,
- through questionnaires,
- through user 'walk-throughs' of the IDDA system, comparing it with other existing systems,
- through user questionnaires comparing the proposed system with existing systems.

These were chosen because a 'user-based' approach to evaluation was more suitable for this project than a 'theory-based' approach. It involves one or more users completing one or more tasks in an appropriate environment, i.e. one that closely resembles the environment for which the system is intended. It also allows the system to be trialed both during prototyping as well as in its final state.

Moreover, it was intended that the techniques used should capture the three types of evaluation data that have been listed as being important, i.e. objective, subjective and cognitive. Sweeney et al (1993) suggest that objective performance data such as time to complete a task, accuracy or error types, can be measured during sessions where the users are asked to complete one or more (bench-mark) tasks. Cognitive data, however, such as the user's understanding of how the system works, can be elicited either by recording the users' verbal comments made concurrently during the interaction or by recording the users' post-hoc comments after using the system. Finally, subjective data such as user opinion or attitude towards the system can be assessed via questionnaires. Subjective data are important as they provide a measure of user acceptability, which is, in fact, a broader concept than usability and likewise as crucial.

Sweeney et al (1993) summarised a variety of data capture methods which could be used for assessing usability and user acceptability. The following techniques are applicable for this thesis:

	Examples of the type of data that can be captured	Observation of user during interaction	Audio-recording of user during interaction	Audio-recording of user's post-hoc comments	User questionnaires	'Walk throughs'
Performance (based on user interaction)	Task times, % completed, Error rates, Duration of time in HELP	√				√
Non-verbal behaviour	Duration and frequency of document usage	√				√
Attitude (User opinion)	Comments, Questionnaires, Discussions, Rating of usability properties	√	√	√	√	√
Cognition (understanding)	Verbal protocols, Answers to questions		√	√		√
Stress	Ratings of anxiety				√	√
Motivation	Enthusiasm, Willingness and Effort		√	√	√	√
Conformance	Comparisons with: - other systems, - design criteria					√

Figure 6.1: Data capture methods for usability and user acceptability factors

A brief description of each of these data capture techniques is given:

Observations:

This involves 'looking over the shoulder' of users whilst they operate the system and noting any problematic aspects of the interaction. Observations can be conducted relatively quickly and they provide first hand feedback of user interaction without requiring the users to curtail their movements at the computer or with co-workers. Sweeney et al (1993) believe that, 'observation can potentially yield as much insight as a lengthy and costly experimental evaluation. There may be additional facets to the interaction which may require attention (such as the use of documentation, the need for pen and paper, supplementary information, etc.) which methods such as automatic monitoring omit'.

There is however no permanent record of the interaction which can be reviewed later, unless the session is video-recorded. Video recording requires additional equipment with the capability of capturing the screen, the input devices, the user's facial expression and the movements of the user, simultaneously. It also restricts the user movements to the camera's field of vision and requires substantial amounts of time for analysing the recording. For example, Berry and Hart (1990) estimated that it takes approximately 10 hours to analyse a one hour tape. Moreover, many users feel unnerved at being filmed and, as Sweeney et al (1993) stress, it is important that the users are aware that it is the system which is being tested and not them.

Consequently, with this project, before the trial started the users were given a brief talk in which the system was described, the tasks that they would carry out were explained and the purpose of the observation was outlined. In this manner, the users knew the aim of the evaluation and understood, in basic terms, the phases of the construction process. It was decided that as minute detail was not required, the manual approach of recording the information would be adopted since it was :

- easier to initiate and run a trial when there was no requirement for specialist equipment or extra space in which to set up the camera,
- more informal and therefore the users would be more relaxed and natural during the trial,
- adequate to capture the details required, i.e. the noting of general user performance, the users' concurrent comments expressing their opinions of the tasks they faced, and any use of extra documentation or pen and paper.

Audio-recording of user's during their interaction with the system:

This can take the form of requesting users to 'think aloud' as they operate the system, thereby recording an explanation of what users are doing and why, or it can involve the recording of users' ad-hoc deliberations and comments as they are using the system. The first choice suffers from the same difficulties as protocol analysis in knowledge acquisition (these were reviewed in Chapter 2). The second method, however, can be used to highlight those areas in the task where users have problems, as well as those instances when users have sudden insights into what is being asked of them. It can also give indications as to what users are attempting to do, e.g. 'I now want to add another record'. As it is not intrusive, i.e. it does not require all actions to be verbally explained, it does not effect the interaction that is being monitored, although it will capture the user's general attitude towards the computer and the program.

Therefore, for this project, a decision was made to tape record the users' ad-hoc comments during their interaction. These could then be reviewed after the session, thereby adding further information to the observation notes. Tape recording was thought to be the best method with which to capture the emphasis and choice of the users' words, thus gaining a realistic impression of the users' views of the system. As explained above, this method is not intrusive and yet can assist in determining the users' general opinion and understanding of the system and the tasks being undertaken. It can also be useful in uncovering the users' overall motivation and enthusiasm for the system. Hence the belief was that this technique would be a good partner to the notes already taken during the observation sessions.

Audio-recording of user's post-hoc comments:

These too should be taped since 'it is surprising how often information not thought to be relevant at one point in time becomes very relevant at a later point in time' (Berry and Hart, 1990). Post-hoc comments provide a useful insight into users' conceptual models of how the system operates and users' opinions of the ease of use, learning, flexibility, etc.

Consequently, during the evaluations for this project, a tape was left recording after the tasks had been completed and users were asked how they felt the trial went and whether they had found any tasks difficult to accomplish. The responses given led to further prompts with the purpose of discovering the users' attitude towards the system and their understanding of the system. From the manner of their responses, the motivation and attitude of the users could be obtained.

This method, therefore, recorded users' views immediately after the trial and these were then compared with the comments that users made during the trial. It was also used to ensure that at least some of the user's opinions were gathered, since a few users may not make any verbal comments whilst they were completing the tasks and to record comments from some trialists and not others would be inappropriate.

Questionnaires:

Questionnaires can be a very effective means of gathering opinions and for summarising viewpoints (Berry and Hart, 1990). They should include both open and closed questions as this mixed composition can cover the domain more effectively. Open questions have no response suggested and people write in their answers in the allocated space. Closed questions, however, provide two or more possible answers and people select the most appropriate response. The choices may be simple alternatives such as 'yes' or 'no' or they may be 'multiple choice', covering various shades of opinion.

In both the audio recording of the users' interaction and the recording of the post-hoc comments, the users themselves drive the commentary. Consequently the information gathered is likely to be different for the individual trialists. It was considered important to also gather more directed views and opinions which could be compared with each other later. Therefore questionnaires were given to the users to complete after their post-hoc comments. (A copy of the questionnaire is in Appendix D).

The questionnaire contains both open and closed questions to prompt the users to express their views in a number of different ways whilst still directing them, when constructing or selecting a response, to focus on specific aspects of the trial and the system. These individual responses to the questionnaire can then be contrasted to gain an overall impression of the users' attitude towards the system.

Questionnaires were also used to acquire the views of the medical consultants on the benefits of a system, such as an IDDA system, to the medical domain. This questionnaire (in Appendix D) was only given to the medical consultants as it was felt they were the only trialists who could make such a judgement. In addition, a questionnaire prompting the medical consultants to compare the constructed IDDA system with a current hand built system was also compiled (in Appendix D). It was anticipated that with these different questionnaires, the possible benefits and the impact of the IDDA system could be reviewed.

'Walk-through':

'Walk-through' allows an evaluator to work through specified tasks using the system. These evaluators could be the developer, at the initial development stage, or trialists attempting to view the system from the perspective of the proposed users. Consequently, the first step is to define the intended user group and identify the key characteristics of the users. There has been the suggestion that this technique is appropriate when new or prototype systems are being evaluated, as there are no users with direct experience (Berry and Hart, 1990).

However, it is also useful when comparing a new system with an existing system. Here evaluators carry out specified tasks with both systems and comment on their approaches, their preferences and opinions. The evaluator could be a frequent user of the existing system but this may introduce bias. It is, therefore, better to select evaluators who know the domain and are aware of the existing system and its facilities, but who have not used that system extensively. This will allow for a more realistic evaluation to be carried out, since a new system can not, in reality, be totally different from an existing system without causing problems with user acceptability, e.g. it is hard to convince all the users that the necessity of re-training to use the new system would be highly beneficial to them.

Two different 'walk-throughs' were used in this project. For most of the trials the users were given a task to construct a specified IDDA system. However, 'walk-through' was also used to allow the medical consultants to compare the resultant IDDA system with a current hand-built system. Notes were recorded during both of these types of 'walk-through' by using the methods explained above.

It was necessary to use this technique as the IDDA end-system was a prototype of a general system and the tools, themselves, were new. Hence there were only a few individuals who had previous experience of using similar systems to the IDDA system and no individuals who had previous experience of utilising the tools. Consequently, specified tasks had to be assigned to the users. Obviously, as the tools are used over a period of time and in different environments, the user base will grow for both the tools and the IDDA systems and further evaluations can be undertaken with these knowledgeable users to determine the subsequent direction in which the two systems should evolve.

In selecting the data capture methods for this project, an attempt was made to adequately cover the usability issues that have been identified as being of interest to this project as well as utilising techniques that would not be onerous to the users or that would require expensive equipment.

There was also the realisation that the selection of evaluators was very important, as it is necessary to have a representative sample of the proposed user group whilst not using individuals who may have strong personal bias for one particular system or another. Users do differ and hence a range of users with different backgrounds and experiences is required.

For this project, initial trials were undertaken by the author and the project supervisor to ensure that the tools and subsequent IDDA end-systems were constructed correctly and that each would complete the tasks expected of them, i.e. to evaluate the 'technical' performance. A subsequent set of trials were then carried out by users who varied both in computing experience and in knowledge of the specialist domain, i.e. to evaluate the effectiveness of the system with the potential user group. They had been selected because of these differences in background and current skills. More details of their characteristics are described in the relevant sections of Appendix C. There was, however, an attempt to create an even distribution between male and female trialists and this was achieved successfully.

Therefore the first two major trials were assessed by two users (the author and supervisor). These were followed by a number of 'real' medical applications undertaken by the author (see table 6.1).

Evaluation	Evaluator	Computer knowledge	Specialist domain knowledge
MOT (i)	Author	Extensive	No
MOT (ii)	Supervisor	Extensive	No
Poll Tax (i)	Author	Extensive	No
Poll Tax (ii)	Supervisor	Extensive	No
Knee Ligament	Author	Extensive	Some
Knee Replacement	Author	Extensive	No
Patella-femoral	Author	Extensive	No

Table 6.1 : Initial trials testing the tools during the construction of different systems

The subsequent trials involved users with different backgrounds (see table 6.2).

Evaluation	Occupation	Computer knowledge	Specialist domain knowledge
Pat-fem. (I) Trialist 1	Researcher in Computing	Extensive	No
Pat-fem. (I) Trialist 2	Physiotherapist	None	Some medical knowledge
Plant trial	Biologist	Some	Yes
Pat-fem (II) Trialist 1	Secretary	None	No
Pat-fem (II) Trialist 2	Secretary	Very limited	No
Pat-fem (III)	Consultant surgeon	None	Yes

Table 6.2 : Relevant background experience of the other trialists

Finally, although the two consultants who compared the IDDA system with a hand-built system had similar medical backgrounds i.e. both were experts in their specialist domains, one was unfamiliar with using a computer whilst the other had previously used a computer to assist him in a few tasks and therefore could be categorised as having 'Limited' computing experience (see table 6.3).

Evaluation: comparisons	Computer knowledge	Specialist domain knowledge
Knee Ligament	None	Yes
Knee Replacement	Limited	Yes

Table 6.3 : The characteristics of the medical consultants comparing the two systems

With regards to the environment in which each trial was to be held, the decision was taken that the best approach would be to use the natural working environment of the trialist, i.e. their offices, and at a time that was convenient to users, i.e. users could select both the time and the day. In this manner, the trial time and place were known and planned in advance and users were aware of the schedule and was relaxed within their surroundings. Hence, it was anticipated that users would be able to concentrate fully upon the tasks required in the trial and not be distracted by external influences or by an unfamiliar environment. During the trial, they had access to an on-line help facility but had no written documentation to read beforehand. They were given a brief talk describing some of the facilities of the tools and either a copy of the questionnaires they were to encode or a sheet detailing the actual data they needed to enter.

Full details describing each trial are in Appendix C. However, following is a brief summary of some of the findings from these trials.

6.10 Summary of the findings from the evaluations

Each of the evaluations undertaken resulted in the successful construction of the intended IDDA end-system, including those carried out by trialists who had neither previously used the tools nor had been involved in their development. The following tables (tables 6.4 through 6.6) briefly highlight a number of the findings from these evaluations:

Evaluation	Computer knowledge	Specialist domain knowledge	Encoded a set questionnaire or just the data or own questionnaire	Time to construct Patella femoral end-system	Details in Appendix C, section?
Pat-fem. (I) Trialist 1	Extensive	No	Set questionnaire	36 mins.	IV
Pat-fem. (I) Trialist 2	None	Limited	Set questionnaire	43 mins	IV
Plant trial	More than average	Yes	Own questionnaire	n/a	V
Pat-fem (II) Trialist 1	None	No	Actual data	35 mins	VI
Pat-fem (II) Trialist 2	Limited	No	Actual data	45 mins	VI
Pat-fem (III)	None	Yes	Own questionnaire	60 mins	VIII

Table 6.4 : Trial details

Evaluation	User understands how the tools work	User found the tools easy to learn to use	User found the tools easy to use generally	User felt confident using the tools	User felt in control of the situation	User's overall attitude	Comparative ranking of user's opinions
Pat-fem. (I) Trialist 1	Yes	Yes	Easy	After first couple of screens	Yes	Very Good	1
Pat-fem. (I) Trialist 2	Mostly	Yes	Easy	Half way through	Mostly	Very Good	5
Plant trial	Yes, though unsure at the start	Yes	Very Easy	After first couple of screens	Yes	Very Good	1
Pat-fem (II) Trialist 1	Reasonable amount	Yes	Easy	Half way through	Mostly	Good	6
Pat-fem (II) Trialist 2	Mostly	Yes	Easy	Near the beginning	Yes	Good	4
Pat-fem (III)	Yes	Yes	Easy	Near the beginning	Yes	Very Good	3

Table 6.5 : Users' opinions

Evaluation	Tools operate at appropriate speed for the user	Dialogue style appropriate, e.g. menus, instructions	Appropriate error messages	Appropriate amount of information given	Appropriate help facility	Data entry method appropriat e	Amount of effort required to carry out task	Comparative ranking of system usability
Pat-fem. (I) Trialist 1	Yes	Very easy to use	Very easy to understand	Yes	Very easy to use and understand	Very easy	Very little	1
Pat-fem. (I) Trialist 2	Yes	Easy to use	Very easy to understand	Yes	Never referrenced	Easy	Some	4
Plant trial	Yes	Easy to use	Sometimes easy/difficult	Yes	Easy to use and understand	Very easy	Very little	2
Pat-fem (II) Trialist 1	Yes	Sometimes easy/difficult	Sometimes easy/difficult	Yes	Never referrenced	Very easy	Some	6
Pat-fem (II) Trialist 2	Yes	Easy to use	Easy to understand	Yes	Easy to use and understand	Very easy	A little	4
Pat-fem (III)	Yes	Easy to use	Easy to understand	Yes	Never referrenced	Easy	Vary little	2

Table 6.6 : System usability and user acceptability

From the Users' opinion summary table (table 6.5), it is noticeable that the trialist who had neither computing knowledge nor knowledge of the specialist domain (Pat-fem II, Trialist 1) selected the lowest categories of all the recorded profiles. A similar picture can be seen with the System usability table (table 6.6), except for the 'Data entry' column. This particular response could help explain the unusual situation where the trialist who recorded the greatest uncertainty actually completed the task of successfully constructing an IDDA system in the fastest time of all the trialists. It is likely that this outcome was due, in the main, to the design decision to use a keyboard as the data entry device. This trialist was a secretary and although she had not used a computer previously, she did have extensive typing experience and therefore this mode of entry suited her background particualrly well.

It is however interesting that when reviewing this lowest recorded profile, the user still: found the tools generally easy to use and learn, felt confident using the tools by half-way through the construction task, and felt, in the main, in control of the task and the computer. Although her understanding of how the tools worked was not detailed, her overall impression and attitude towards the tools was good. It is sometimes difficult to be enthusiastic towards something if someone can not see the benefits themself, i.e. if a person's understanding of what a system does and how it will assist them in their job, are not clear. This could be the situation here as it is noticeable that the other secretarial trialist also had a similar overall attitude towards the tools even though her recorded profile was better (Pat-fem. (II) Trialist 2). All the other trialists could either understand the purpose of the tools from a computing perspective or could relate the facilities that the IDDA system would offer to their own jobs and thus they could understand the benefits of using the tools to build such a system.

The only other trialist who had no prior computing experience was an expert within the specialist domain. It appears that the familiarity of the field assisted him during the interaction since his recorded opinions of the usability and acceptability factors were better than Pat-fem. (II) Trialist 1, as was his overall attitude towards the tools.

From the comparative rankings, it can be seen that a trialist's view of the tools is influenced by their computing knowledge. However, if a trialist has in-depth knowledge of the specialist domain, this seems to assist a user. Therefore, as the tools have been designed to allow experts within a specialist field, who may well be naive computer users, to build their own IDDA end-system, the indications from these evaluations are that such operators will be able to successfully build such a system and that they will feel confident and/or happy using the tools early in the construction phase.

Comparing the various times (see table 6.4) taken to construct the Patella femoral end-system, it is noticeable that the medical consultant took the longest time. This could have been due to his lack of computing knowledge and/or his lack of typing and keyboard skills. It could also have been influenced by his knowledge of the domain as he was attempting to model the responses to the questionnaires himself rather than accepting the printed questionnaires and data per se. Even so, one hour to build an end-system from a series of questionnaires is certainly not excessive.

In fact it is less time-consuming than involving a third-party to hand build the system, as the builder would need the requirements specified and explained before any construction could begin. Moreover, using the tools, the construction of the system was completed in the recorded time whereas in the other scenario the system would still have to be designed and built. Furthermore, the consultant would be reliant upon the third party and whatever time that person may or may not have available. Generally, building a basic database system for gathering data would require a number of days and a system to accommodate naive computer users would need a timescale in the order of weeks. Any advanced features such as user-specified statistical analyses would require substantially longer. Hence, domain specialists who use the tools to build an IDDA system in an hour will be saving large amounts of their time, their unit's money, and, the time-lag before the data collection involved in the investigation can begin.

It must also be remembered that the hour recorded in the evaluation was in fact the first time the consultant had used the tools. He stated, in the post-hoc comments, that he felt happier and more confident as the trial had progressed and he thought that 'next time' he could complete the task faster and with more ease than during this first attempt. Therefore these expected gains, in terms of time and money from utilising the tools rather than employing a third party, seem to be achievable.

With regards to the issue of who should operate the tools, if an operator can spot and correct errors during the building of an end-system then this ability is more advantageous than being able to construct a system 15-20 minutes faster but having no understanding as to whether the system is 'correct' or not (see Appendix C (VI)). Consequently, there will be the recommendation that the operator of the tools should have a fairly in-depth understanding of the specialist domain, i.e. the role of the secretarial support staff will be restricted to the production of the ASCII text file of the questionnaires and the help text whereas the tools will be operated by either a domain assistant or an expert. This recommendation is also based on the realisation that the operators who benefit directly from the use of the IDDA end-system, have the best attitude towards the tools and thus are more enthusiastic and supportive than those users who perceive they do not gain in their tasks from the use of an IDDA system.

When reviewing the users' opinions of the interface, the input method selected seems to have been appropriate as does the dialogue style and the amount of information displayed. The low utilisation of the Help facility seems to indicate that the tools were, in general, relatively straight-forward to use. However, when referenced, the help was classified as easy to understand and follow. The users' views of the error messages varied considerably and do not seem to relate to the users' backgrounds. Even so, the error messages were still classified as only being 'sometimes difficult to understand', but they will need to be monitored during subsequent long-term trials to determine if the appropriate explanation level is being addressed.

The amount of effort required by the users to interact with the tools to build an IDDA system again seems appropriate (see table 6.6), with only the trialist who had neither computing knowledge nor knowledge of the domain finding the tasks somewhat taxing. As for producing the assessment questionnaires from which the IDDA system is built, this task appears to be the longest and most trying section in the construction process (see Appendix C (V) and Appendix C (VIII)). However, the time taken is no longer than would be required for producing the questionnaires to be used during a manual assessment exercise. Hence, the tools are not imposing any further overhead in the production of the questionnaires than would normally exist. Moreover, as explained in Appendix C (V) (see page C-52), the time spent at this stage is in fact beneficial, as it will result in a better planned and more justifiable study being conducted.

The conversion of the paper assessment questionnaires into ASCII text computer files via a word-processor is reviewed in Appendix C (I) but the indications are that this step does not cause problems. Therefore it would appear that the contribution required from the users to construct an IDDA system through utilising the tools is neither excessive nor onerous and thus user acceptability and support for the tools should not be adversely effected.

A number of other issues were noted during the post-hoc comments:

a) one trialist, who was computer experienced, did not like the colours selected for the interface screens of the end-system and would have liked the opportunity to have been able to choose his own. No other evaluator commented on the colours. The current colours of: dark green background, yellow text with white system text, i.e. for function labels and error messages; red-brown answer boxes with white entry text, were chosen, after studying the literature, for their restive qualities and the highlighting of important information. In addition, these colours have already been used in a number of systems previously hand-built for the medical domain. The operators of those systems have not reported having any problems with these colours. There is also the question as to whether naive computer users could select reasonable colour combinations prior to building their own systems. Moreover, if such a facility were provided, what would be the delay to the construction process?, and, could it cause confusion as well as disputes between intended users with different opinions over which colour combinations to select?

   As only one person has commented on this, there will be no changes made at present. However, as more end-systems go into operation, they will be monitored to determine if difficulties develop after prolonged use.

b) another trialist would have liked the function key labels in both the tools and the end-system to have been placed at the bottom of the screen rather than the top. Again, from previous research of the literature, the indications were that the top of the screen was the best position for displaying labels for extra facilities as this position appears to better capture the user's attention. Once more, as only one evaluator has commented on this feature, a review will be undertaken after a number of IDDA systems have been in daily operation for a period of time.

c) two evaluators commented on the ESC key (the undo facility) in the tools which returned them back to specifying the answer type for the questionnaire's question rather than just back one definition question. Other evaluators seemed happy with the current facility since they often made a couple of mistakes before realising the error. As explained in Appendix C (IV), little or no time will be lost with the current method. In addition, it is likely that the reason for this annoyance has now in fact been removed when the extra confirmation questions were taken out, (as explained in Appendix C (VI)). There is evidence that this problem has been solved since no subsequent evaluators reported a problem with the operation of the ESC key.

d) the issue of the second screen was raised by one trialist and could have been prompted from the incorrect positioning of the screen (see Appendix C (V)). All the other evaluators reported having no problems using the second monitor. However, as discussed later in Appendix C (VII), a facility to run the tools, and the basic data entry section of the IDDA end-system, without a second monitor should be considered.

Two possible additions mentioned during the evaluations were:

a) the ability for constructing a 'glance' chart, which is a time-series table (see Appendix C (VII))

b) a printing facility (see Appendix C (VIII))

Both are likely to be added in any update of the tools. However, whether a) will be just an add-on facility that can be requested or a standard part of the tools, will be determined from further discussions with medical consultants.

In addition, the following changes are also likely to be made:

a) the removal of tables and list facilities from the tools to avoid causing confusion;

b) the production of further tools to help maintain the constructed IDDA end-system.

Consequently, the summary above briefly reviews the comments and observations made whilst the evaluators carried out the task of building an IDDA end-system. All the trialists successfully managed to construct their system, even though they varied in their knowledge of computers and in the specialist domain. Their reports were all favourable and, in general, they appear to have found the computer tools easy to use, easy to learn, easy to understand and straight-forward and logical so that they quickly felt confident using them. Moreover, the responses recorded in the System usability table (see table 6.6) indicates that many of the design decisions taken were in fact appropriate for the intended environment, task and user group. Obviously reviews of users' opinions regarding system usability will need to be undertaken after the tools and IDDA end-system have been in prolonged use and in a variety of different real-life environments. However, the results from these evaluations are certainly promising.

One of the most interesting measurements was finding the length of time it takes for a user to complete the task of building an IDDA system. It is evident from these trials that using the tools to construct such a system will provide substantial savings in both time and money. Consequently, utilising the tools can lead to real-term benefits for the users.

With regards to the two evaluations involving the comparison of a generated IDDA end-system against an end-system that was hand-built (see Appendix C (VII) and C (VIII)), the following findings are worth noting.

Both the medical consultants, who were shown the generated IDDA systems, were impressed with the facilities offered and the ease by which these system could be operated. They were particularly interested in the analysis and review capabilities of these systems since it was in this area that their own systems were found to be lacking. Appendix C (VII) and (VIII) contain the descriptions the doctors gave of a number of short-comings in their current systems and how they perceived the demonstrated IDDA system could overcome some of these difficulties.

In their opinion, the generated IDDA end-system was better designed and equipped to deal with the requirements of doctors working in specialist medical domains than their previous systems. They believe that using an IDDA system will help to ensure that all the relevant information is gathered during a patient assessment session whilst also assisting in the easy retrieval, browsing and manipulation of data during an investigation.

Reviewing and analysing data is currently the most frustrating and arduous tasks of clinical trials. Doctors have to collate from the original data sheets, the information needed and then re-enter this data into a statistical package in the correct format, to undertake the desired analysis. Different analyses may require different data sets and so the process needs to be repeated. This increases the likelihood of human error and inaccuracies in the translation of the data. If, however, the doctors have been awarded a grant to carry out the study, they are more likely to dispatch the collected data sheets to a statistician and wait for the results of the requested analyses to be returned. Neither of these two scenarios is ideal and therefore the possibility of having a system with inbuilt data collection and statistical capabilities was very appealing. Especially as this system enables the doctors to carry out tasks which previously they could not do themselves, or which were so time-consuming and tedious that the likelihood of error was high.

This ability to undertake user-specified analyses on past-cases permits investigations to be carried out into the decision-making processes in the specialist field as well as reviewing the various patient outcomes which occurred. Thus, evidence can be gathered to support or refute theories and established beliefs. Moreover, such investigations could result in explicit management strategies being developed, thereby reducing the current reliance on ad-hoc selections of methods.

The consultants also hoped that audits on present practices and tests, will reveal which techniques are of use and which are of little value. Thus guidelines will begin to emerge as to which procedures should be used during data gathering and which should be avoided. Hence a more formalised, standardised field may evolve, which could lead to better health care being provided by, for example, enabling savings in time, money and discomfort to patients.

In addition, the ability that the tools give the doctors to quickly construct appropriate systems for studies, permit the consultants to feel in control by providing them with the opportunity to respond to the demands of an environment which is dynamic and constantly evolving. This ability for systems to be built by the doctors themselves will create further savings in both time and money since it will not be necessary to employ an external contractor everytime a new study is initiated. This will, the consultants think, encourage more research to be undertaken since now the cost for starting an investigation will be minimal. Moreover, the review and analysis capabilities of the IDDA end-system enables the doctors to carry out their own analyses and reviews rather than involving statisticians. Thus, more investigations can be undertaken more quickly and cheaply than before, resulting in more information being gleaned from the collected data. Hence, the idea of initiating a trial becomes more appealing in the first place.

The final benefit listed was the belief that an IDDA system would also fit into the working practices of the medical unit with minimal disruption to either staff or patients. Hence, the acceptability of the IDDA system to the medical profession is not affected by the fears and concerns often expressed regarding the perceived threat that the use of computer systems will have on the patient-physician relationship and/or on the current working practices of a unit.

All of these benefits, the consultants believe, could be realised by utilising an IDDA end-system. Hence, such a system would be very useful in medical investigations and studies. Consequently, the responses from the medical evaluators have been that they would readily accept the assistance offered by an IDDA end-system (and, in fact, have already requested to use the tools for future studies) for they do believe that clinical care can be improved by utilising such a system.

It must be remembered that most short-term evaluations have limitations and the evaluations undertaken for this research have been subject to similar restrictions. The number of trialists is low because the time involved for domain experts to conceive and produce a trial with which to test the tools would be too great to interest such parties in any evaluation. Therefore one particular trial was used as the basis for the evaluations (the Patella femoral). This technique did, however, enable comparisons to be undertaken between different trialists.

The evaluations were very much user-based rather than formally-based as this technique enables users' attitudes towards the tools and IDDA systems to be revealed. As already stressed, if users did not find the tools easy to use or were not willing to accept an IDDA system and incorporate it into their working practices then any benefits derived from utilising these systems would be lost. Therefore determining users' views and attitudes was the main concern during these initial evaluations.

The Evaluation questionnaire, on reflection, could have been more directed and precise. Trialists did write comments next to answer lists and therefore this indicates that the questions and answers could have been improved. However, the audio-recording of users during their interaction with the tools and their post-hoc comments did provide very useful information and did enable clarification of their responses to the questionnaire to be obtained.

However all these factors will have influenced the conclusions drawn. As explained throughout Section 6.9 and Section 6.10, long-term evaluations of the tools and IDDA systems will be undertaken and will involve a variety of domain experts within different specialisms, operating within different working environments. It is anticipated that evidence from these extended evaluations will provide more detailed results and will enable a more critical review of both the tools and IDDA end-system to be undertaken.

Chapter 7
Summary

7.1 Introduction

This project investigated devising a methodology to enable naive computer users, who are specialists in their own domains, to construct an appropriate IDDA end-system for their studies. The composition of the IDDA system was determined by reviewing the requirements of the proposed user group, the tasks to be undertaken and the environment in which the system was to operate. These investigations considered the types of computer system which have already been developed for specialist medical fields and assessed the difficulties and benefits of such systems. It became evident through these investigations that not only was there a requirement for the design of an appropriate IDDA end-system, but that similar systems for different studies would need to be constructed frequently and hence the domain experts needed the ability to construct these systems themselves. Therefore, a suite of computer-based tools was devised which would acquire, from the user, the necessary information to build the relevant IDDA system for the study to be undertaken. The medical domain was selected as the initial trial ground for the methodology as it contained many specialist fields which were unformalised and non-standardised. The indications from the initial evaluations reported that such an approach would not only be successful but would also be acceptable and extremely useful to the intended users.

This chapter briefly reviews some of the issues discussed in the previous chapters and then summaries the likely benefits which could be gained from utilising the methodology proposed. First, however, a few of the problems of current medical support systems are highlighted before the importance of empirical investigations in specialist fields is outlined. The tools and the IDDA system are then described as well as the reasons for selecting the development methods used. Areas of further research are indicated before the advantages of both the proposed methodology and general computer support in medicine are briefly summarised.

7.2 A few problems of present medical support systems

Currently, there is a noticeable scarcity of computer systems to support decision-making within medical specialist fields and yet, as Chapter 1 illustrates, there is an urgent need for assistance to be given to consultants. A few specialist domains have attempted to formalise their diagnostic strategies and have then tried to automate these processes. They have achieved some limited success. This has occurred primarily within fields which have difficulty in identifying the precise problem or the severity of the complaint but which do have predetermined treatment pathways once identification has been made, i.e. in fields which are relatively well formalised and which have agreed standard investigative tests and treatment procedures. Langlotz et al, (1990) agrees, stating that they discovered that most applications concentrate on 'the treatment of patients for whom the course of disease and the response to therapy are relatively stereotypical'. However, many of these original systems, including those classified as 'successful' implementations, have now been abandoned by their intended users. Other systems never managed to progress out of the research environment in which they were developed, whilst still more never evolved past the prototyping stage. A number of reasons have been suggested for this evident lack of success and for the general lack of acceptance of computer systems by the medical profession. These include:

- the over emphasis on AI techniques and the exaggerated claims concerning the capabilities of such systems (Andriole, 1985),
- the requirement for a substantial commitment of time and money to the design and implementation of the AI programs and the necessary knowledge base (Hart, 1982),
- the lack of demand from the hospital community for computer-based decision-making due to (Kunz, 1984; Shumway et al, 1990) :
  - limited convincing evidence that computer decision support provides more effective or more economical care,
  - the reluctance of health professionals to change habits,
  - the absence of computer training in medical schools,
  - the rigidity and complexity of computer systems,
  - the limited time and the high stress levels with which physicians have to work,
  - the pride of the physicians and the perceived threat of computer-systems to their domain,
  - the development of systems which do not exceed health professionals' own capabilities,
  - the production of systems that are not readily transferable from one location to another.

Difficulties, such as these, have prevented commercial vendors becoming interested in developing products for specialist medical fields and hence the main development work has occurred within the research environment, which is itself limiting and limited.

One major problem of developing such systems has been that to understand the medical decision-making process, the important characteristics of the patient and disease must be known. These specifications must also include the relevant weighting or 'value' of each characteristic within the decision-making process and its influence or impact on others. However, as has been reported widely and as summarised in Chapter 2, this knowledge is extremely hard to acquire from any source and is equally difficult to model in an appropriate fashion. Yet, if the systems are to manipulate information and to attempt to derive the same conclusions as the human experts, they must be provided with the same knowledge on which to base those decisions.

Aitkenhead and Slack (1990) have identified humans as being animate organisms with a biological basis and an evolutionary and cultural history. Moreover, they are social animals, interacting each other, with the environment, and with themselves. For example, Thomson (1966) described a lone thinker at work, 'he may have an intuition that a particular conclusion is the one he is striving for and then construct a chain of inferences to deduce this conclusion from well-established premises. He may ask himself specific questions and give answers in the light of facts or theories known and remembered by him. He may argue the case - taking first one and then the opposite side in the issue; he may cross-question himself - playing the roles of 'witness' and 'counsel'. During all this, he may work it all out in his head, or he may utter his thoughts out aloud to himself or he may jot them down in writing. He may consult books, papers, and mathematical tables in order to complete an inference.'

These aspects of behaviour, however, have been ignored during the construction of knowledge-based systems. Instead, humans have been conceived as being pure intellect, communicating with one another in a logical dialogue, perceiving, remembering, thinking where appropriate, and reasoning a way through the well-formed problems that are encountered in a day (Aitkenhead and Slack, 1990; Akman and Ten Hagen, 1989). This scenario is rarely the case and therefore this conception does not fit actual human behaviour. Humans do not work within tightly constrained environments. Also they often relate knowledge and information from other episodes within their experiences and thereby have 'the ability to say "Wait a minute, this isn't right. I'm getting the hell out of here. I'm not going to try and see if it's this rule or that rule". It's precisely that ability to back off from a domain as initially perceived and treat it as part of a larger domain that distinguishes the sort of narrowly focused expert systems that we now use from what you might call a real expert who carries with him/her that whole ability' (Davis, 1989b).

Although computer systems can be built which can gradually refine their performance, they do still require a basic model from which to start and the correctness and completeness of this model will directly relate to the accuracy of the results gained. However, the structure and the functioning of the human brain are still unknown and yet systems are being, and have been, built to imitate the human decision-making processes. Moreover, knowledge-based decision support systems have not currently been based on a general, formal theory of decision-making nor even on a clear statement of what a decision is (Fox and Krause, 1992). Without any basis in formal theory of decision-making, there are no clear criteria available for judging the soundness of a program, or predicting its performance, particularly in adverse conditions. Therefore, this situation is of grave concern especially when these systems are implemented and used in tasks which directly, or indirectly, effect other people.

Within some specialist medical fields, the diagnosis is relatively straightforward. However, the selection of a treatment or therapy strategy is not. Such fields tend to be unformalised with no agreed standards regarding which tests or procedures should be carried out to collect data and information. Thus with no established case history database to investigate and analyse, decision-makers are left to continually rely upon their own experiences and personal preferences when selecting a treatment strategy for a particular patient.

Although there will be variability in how individual patients respond to equivalent treatment strategies, the structure of the body is relatively invariant amongst a large group of patients. Uncertainty about the behaviour of a specific patient implies that uncertainty must exist about the consequences of carrying out the different treatment actions, especially when there is such difficulty in defining 'success'. As Mouradain (1990) asks, 'what exactly is a 'successful' surgery - when the patient would do it all over again? When they return to work? Or when the surgeon thinks the surgery is a success?'.

Currently, decisions regarding treatment strategy within such fields are based on experts' judgements and their own personal experiences and therefore any system modelling these unformalised decision-making processes would have to be based primarily on subjective data. Human judgement and thought is known to be profoundly influenced by a variety of biases and some of these arise from that very specialised knowledge itself which characterises the expert. For example, instances of large classes are recalled better and more rapidly than instances of small classes; likely occurrences are easier to imagine than unlikely ones; and associations between events are strengthened when the events occur frequently (Tversky and Kahnemann, 1990). Therefore, although humans have the ability to estimate the number within a class, the likelihood of an event, or the frequency of co-occurrence, by the ease with which the relevant mental operations of retrieval, construction, or association can be performed, these estimation procedures produce systematic errors (also refer to Chapter 5).

In addition to these types of error, there is the problem that people tend to be more confident about the accuracy of their guesses than is warranted, e.g. as Fischhoff et al, (1977) found among college students who estimated the frequency of a variety of causes of death in America. There is also evidence that many non-occurrences of events are ignored in favour of those that confirm their hypotheses (Einhorn and Hogarth, 1981). Moreover, humans tend to be misled by the possible alternatives because they focus on irrelevant features of the problem, or because they are guided by faulty expectations and biases that favour certain hypotheses (Elstein et al, 1978). Therefore, as Fieschi (1990) rightly states, 'no doctor responsible for the life and health of his patients can readily accept proposals from a system based largely on the subjective views of other people'.

It is also common for too much data to be collected in the belief that decisions will be improved, when in fact Koran (1975) has shown that this extra data overloads the cognitive capacity to interpret the data correctly. De Dombal (1978) agrees, explaining that the 'reason why doctors from time to time make erroneous diagnoses is simply because they are totally unable to handle the volume of data which they elicit from patients'.

This particular problem of information overload has grown substantially over recent years as more and more tests have become available and the quantity of research literature has increased dramatically. Yet the ability to be able to recount all the known prevalences of a disease and the outcomes of the previous cases when arriving at a diagnosis or a treatment strategy, would be highly advantageous. As Collste (1992) explains, 'the ideal judgement, the one made at the critical level, presupposes impartiality and total knowledge of the situation and the consequences of the different alternatives. The action that leads to the satisfaction of as many preferences as possible is the one that should be chosen'. However, the mere quantity of data requiring to be stored and sorted and the underlying problems of human memory in terms of recall, associations, weightings and biases prevent this from happening (refer to Chapter 5, sections 5.5 and 5.6). Consequently judgements are being made under non-optimal conditions, thus leading to the occurrence of errors and mistakes.

Therefore, Aitkenhead and Slack (1990) believe the moral is that, 'if your goal is to use a computer to perform some function as intelligently as possible, the best solution may not be to imitate expert human beings.' In fact, rather than having a goal of simply mimicking a human expert, a better aim would be to assist experts in the tasks they find tiresome or difficult but which fall within the capabilities of current computer technology, and thus allow experts to concentrate on those aspects of the decision-making process in which humans excel, e.g. hypothesis generation.

Evans (1989) agrees, believing that the way forward is in the development of techniques for the interactive design of decision aids. These database systems could then be used to assist in the de-biasing of decision-making. For example, by replacing otherwise biased intuitive judgements with accurate computational and database search methods, or, by the design of user interfaces which ensure attention is drawn to the relevant information and the prompting for forgotten pieces of information, or, by assisting the user in structuring the problem and thereby supporting the user during the selection and the application of decision theory methods. Hence, there needs to be a move away from personal preferences and subjective decision-making to a more rigorous method of problem-solving. Computers can assist in this process and can bring substantial benefits. They are indispensable tools that can help to analyse, co-ordinate, store, retrieve and compute as well as compile, at rapid rates, the vast amounts of information that becomes the knowledge-base of any clinical speciality (Kleinmuntz, 1984).

However, a survey undertaken by Millington et al (1991) highlighted the lack of computing experience and cost as factors influencing the low utilisation of computers within healthcare. Computer training will have to become an integral part of medical education. Without it, medical personnel are unable to identify and define the capabilities they require from a system. This results in end-systems being developed which have limited usability and applicability to the task and/or domain in which it is to operate. If, at the same time, there was an increase in the reusability and/or portability of applications or software components, there would exist the necessary additional impetus to justify the initial financial outlay for a system. For as the needs of the medical domain which is so dominated by data and information increases, the requirements for quick and efficient production of applications and the effective reusability of software will continue to grow. Hence, as Millington et al (1991) agree, one of the foremost demands will be for automated methods to be used during the application development process.

Therefore the task for developers will be to provide a mechanism by which experts acquire the computer systems they want, i.e. systems which undertake the tasks experts need in a manner which is sensitive to the demands of the working environment and the level of the operator's computing knowledge. As experts are aware of all of these criteria, the most appropriate methodology to follow would be to supply experts with tools which enable them to build their systems as and when they need them. However to build such tools, a general understanding must exist of the working environment, the tasks which are to be undertaken and the people who would be involved at all the various stages.

7.3 Clinical studies

Within specialist domains, the type of studies undertaken generally follow the empirical scientific approach, i.e. collect data prior to an event, carry out a specified event, record data after the event, then compare the two sets of data and determine the effect of the event. Consequently, the tasks involved in such a study include: the recording of the data at each of the stages, the analysis of this data, using a variety of statistical techniques, determining both the short- and the long-term effect of a specified event.

By computerising the first two tasks, benefits could be achieved by: reducing the errors in the stored data, enabling specific records to be found and reviewed more rapidly, increasing the speed of data analysis, improving the accuracy of the results obtained, and, providing the ability to undertake more complex analyses than could previously be carried out by hand. Therefore the number of investigations and queries on the collected data can be increased as well as these being more detailed and more accurate. This may result in further information and domain knowledge being uncovered.

To enable such studies to be undertaken, complete data sets need to be gathered in an orderly and standardised manner. For example, by using pre-defined questionnaires. Clinical decision-support systems need to deal with medical data about patients. Checklists, such as standard medical questionnaires, would ensure completeness of the gathered information and permit the analysis of the data to investigate the underlying decision-making processes that had been used. Therefore, when attempting to uncover new knowledge within a domain or confirm currently held theories, the importance of collecting data through the use of standard techniques can not be over stated. Clare (1976) agrees, 'when the three stages of the diagnostic process (data accumulation, data interpretation, and data categorisation) are approached in a rational and competent manner, the results in terms of diagnostic agreement and all that follows it compare favourably.'

There is evidence to suggest that the impact of introducing standardised assessments in areas where they currently do not exist will be minimal, for as Rector (1989) points out, 'most consultations follow a stereotyped script, [though] few systems take full advantage of this fact'. De Dombal (1988), a qualified medical doctor, and McDonald (1976) also concluded from their studies that the utilisation of standardised assessments would in fact be highly beneficial as they would force doctors to adopt the old fashioned approach when taking patient case-histories, i.e. talking to the patient carefully; listening to what the patient has to say; defining the terms used by the patient faithfully, and, thus, recording a thorough case-history.

Moreover, these studies have revealed that doctors themselves like using pre-defined structured approaches as they are working in an environment in which external interruptions are frequent and where there are continual pressures being placed on their time. Studies of medical auditing (Donabedian, 1980) have shown that doctors often omit simple checks and make simple errors despite adequate knowledge, due to interruptions or tiredness or haste. McDonald (1976) concluded that there seemed to be no association between the level of training or the perceived skill of doctors and the frequency with which they made routine mistakes. Mistakes which McDonald believes can be easily audited and probably avoided with the help of a relatively simple computer system.

A further benefit of adopting standardised assessments would be the initiation of a move towards doctors using common terminology and procedures (de Dombal, 1988). This would reduce confusion, misunderstanding and the wasting of valuable time, money and effort in carrying out tests and examinations that are in fact inappropriate or inaccurate for the particular situation or scenario under review.

In fact, summative studies, i.e. those investigations that attempt to draw together a number of findings and explain why different results have occurred, have been severely hampered by this lack of standardisation within the medical domain. For example, if the findings of different groups are contradictory, further indecision and uncertainty are injected into the decision-making process when the reason for the lack of similarity in results could well be due to the inconsistencies in the terms used or the operational procedures undertaken rather than to any differences existing between the items under investigation. This is because any findings or conclusions drawn from investigations within unformalised fields will naturally be strongly influenced by the tests and procedures undertaken in the unit where the study was performed. Therefore other groups within this specialist field must relate those documented tests to their own. The overall result is that studies originally undertaken in one unit are duplicated in numerous other units in an attempt to verify and relate the findings and conclusions to the different settings.

With the transferability of computer systems being limited due to the lack of standards, individual systems must be built for each environment and specialist. Currently, as there is a lack of available computer-based tools to construct the required systems quickly and easily, valuable money and time are wasted. Moreover, all this effort is being expended with little or no advancement of the specialist field itself. Hence, a domain which has no or few standards restricts the ability for meaningful dialogues and discussions between workers within a field as well as preventing the ability to initiate collaborative studies and investigations.

Consequently, there is an urgent need to develop more formalised, standardised fields, without which:

- the advancement of knowledge within the domain will be severely restricted,
- time, effort and money will be expended needlessly,
- in medical fields, pain and suffering will be caused unduly.

One approach is to analyse and review the data collected from the different investigative procedures. In this manner, the procedures and practices of the individual practitioners can be audited and discussions can be initiated to determine the applicability of each of the different methods within various scenarios. These studies could therefore lead to the establishment of an agreed specialist terminology and the evolution of recommendations regarding procedures to follow when a particular problem presents itself.

It is only from the analysis of the collected data and the interpretation of the results that suggestions emerge for both changes in the current practices and the direction of future investigations. This is one of the major drawbacks of running paper-based clinical trials, namely that after collecting sufficient data, the even more tedious and lengthy process of data analysis begins. Often this is restricted to only a few pre-selected queries because of the time-consuming task of collating and reviewing the data by hand. Hence, there is little opportunity to experiment or extend the investigation further than those originally planned queries, which results in the loss of much of the potential for discovering knowledge and for acquiring evidence to support or refute previous decision paths and theories.

The use of statistical approaches to examine the data removes many of the influences of personal preferences and biases from the review process, e.g. the investigations will be conducted on the data collected in previous case-histories rather than from discussions of experts attempting to consider, weigh and justify the various outcomes remembered from their own knowledge-bases and experiences. If the data is held within a computerised database instead of in paper files, there is the possibility of quickly relating any item of data with any other. This enables more thorough data analysis to be undertaken, which increases the likelihood of uncovering useful and unexpected information and, thus, gaining further knowledge and insights into the domain.

Obviously, the validity of any results, whether calculated by hand or machine, is dependent upon both the quality of the data used, i.e. following the old dictum 'garbage in, garbage out', and the selection of analysis to be undertaken. Altman (1991) and Gore and Altman (1990) have reported extensively both on the problems associated with designing and running medical trials and on the analysis of the data and information collected and the interpretation of the results. The lack of training of medical personnel in statistics and its increased usage in medical research, have both contributed to the growth in the occurrence of inaccurate or incorrect results and conclusions being published. As yet, there is no computer system which can guard against poorly designed and badly run clinical trials. However, with an automated system, the quality of data 'entered' can be checked to higher degree than in a paper-based approach, e.g. a computer system can: insist on answers to questions before allowing a user to progress, prompt for the next question and thus prevent the 'skipping' of questions, check that an answer given is within the permitted range, check that an answer is the correct type of response for the question, etc. Moreover, if the required statistical analyses are also computerised, there can be increased confidence in the accuracy of results, e.g. no records have been 'overlooked' or included erroneously, no data items have been 'missed' or transcribed inaccurately, no miscalculations, etc. This, of course, is no benefit if the techniques selected for the analysis are inappropriate for the data or the hypothesis being tested.

One method that is used to attempt to ensure the quality of published findings has been the insistence of some refereed journals that the data, the results and the details of the experimental method used in an investigation be reviewed and confirmed by academic peers or by statisticians (Parasaye and Chignell, 1993; Altman, 1991; Blascovich, 1987). The availability of these extra details of the investigation enable other researchers to interpret the results more accurately and allows them to relate the findings to their own patient population in an appropriate manner. For as Gage (1993a) observes, 'virtually all clinical studies have some methodological weakness, and if clinicians used only studies that perfectly matched their clinical questions, they would rarely find even one such study'.

Furthermore, certain data items may well be able to be shared amongst researchers in different groups and at different institutions - so long as safeguards are taken to ensure that the validity of the data is not affected (see Altman, 1991). Hence, the expensive and slow process of data collection can be eased considerably and the productive aspect of a clinical study, the data analysis, can be initiated earlier. Therefore, the ability to quickly provide requested items of data for inspection or distribution is a further advantage of storing the information in a computerised form.

Unfortunately, many clinical investigators have failed to learn how to use statistical methods, to write programs or to manipulate computers. As a result, they have had to consult specialists in these domains. At present, Akazawa et al (1991) (members of a hospital's statistics unit) are asked by clinical researchers about 200 times a year to extract data sets for statistical analysis from the hospital's case history database. In the future, as the number of users increases, it will be difficult for the unit to fully answer clinicians' requests. This fact has now convinced Akazawa et al that there is a necessity to develop a versatile and flexible statistical analysis system for clinicians. The clinicians requested that such a system should have the following facilities:

1) to perform statistical processing in an interactive mode. Statistical processing should be consistently carried out in a menu-driven, question-and-answer manner. As a result, clinical investigators would not have to enter various commands or write programs.
2) to have self-consistent and extensive HELP functions. Every screen should include detailed program instructions. In the case of erroneous manipulation, the system should show not only error messages, but also an adequate method of recovery.
3) to perform various analyses. The system should be able to perform interactively various analyses used in medical fields.
4) to have access to large archival databases and to extract easily clinical data for statistical analysis.

However, the clinicians are also likely to have limited computing knowledge and lack in-depth or extensive experience of using computers. They could therefore be classified as naive and infrequent users. Nevertheless, they are specialists within their own fields and as such do not want to feel that either their status or their position is being threatened by their ignorance of computing. There is also the problem that the specialists are very busy and are often interrupted when only having partially completed a task. Yet it is the medical specialists who tend to be the initiators of any clinical study undertaken within their domain and they are generally impatient to start an investigation once it has been designed. Therefore, medical specialists want an appropriate computer system to be made available immediately the study is ready to begin. Consequently, the ability to quickly, cheaply and easily construct an appropriate database system for any new study becomes a very important factor. The end-system must however fit into the daily working practices of the unit as well as the physical working environment. As domain specialists know these constraining factors and have designed the clinical study, the best solution would be to establish a suite of computer-based tools which would allow specialists to develop their own database system as and when required.

7.4 The tools and the IDDA end-system

Further reviews were undertaken to determine in which tasks medical users required assistance. It emerged that currently these all revolve around data collected during studies held within the specialist field. Consequently, the IDDA system, which is constructed through the methodology described in this research, is based on analysing and reviewing data gathered during investigations which have been initiated by a user. Thus, there is no requirement for experts to articulate their knowledge or the associations between items explicitly nor are they asked to attach values of importance to the pieces of information. They are not criticised by the machine, or a non-specialist, for not being able to fully justify their actions nor are they queried at length for not being able to explain in enough depth or detail the pathways or knowledge used when they make decisions. The unstructured and arbitrary nature of human decision-making, which exists especially in unformalised fields, is therefore not laid open for scrutiny or questioning nor is confidence in being able to make rational decisions undermined.

Consequently, rather than the knowledge-acquisition difficulties described in Chapter 2, factors such as:

- the characteristics of the intended users, their current skills and the daily demands placed on the user, by both the task and the environment,
- the tasks to be undertaken by the computer and the necessary interaction with those tasks which are not to be computerised,
- the general environment in which the system is to operate,

become extremely important and, thus, have influenced the design of both the tools and the IDDA end-system (see Chapter 3 and Chapter 4).

In Chapter 2, the question of whether experts can design their own system was discussed and the conclusion was that by using the suggested methodology, experts could successfully design an appropriate IDDA system for their specialist fields. The next question is whether an inexperienced computer user can interact with a set of computer-based tools to develop the designed system. There is some evidence to support the view that they can, for example, Tuhrim and Reggia (1986) discovered that if physicians are given a suitable software environment, they can directly implement their own small medical decision system.

Therefore, a decision was taken to design and develop a suite of computerised tools to provide the crucial support and assistance to specialists during the task of constructing their IDDA systems. By providing domain experts with the ability to design their own systems, they can ensure that the system is oriented towards the working practices of the specialist unit, for example, in the terminology that is used, the questions that are asked, the sequence of those questions, the help and explanation that is incorporated into the end-system, etc. Thus, when operating the IDDA system, the users would be utilising their previous knowledge of the domain. Moreover, the system will easily integrate into the working environment and will be able to undertake the tasks required since the expert will have designed it specifically for the domain and the investigation about to be undertaken.

The tasks involved in constructing and running the IDDA system were divided into sections to be allocated to the most appropriate person within the specialist unit. In this manner, the various skills and experiences of the different individuals in the group are utilised and the requirement for the adaptation of known skills, or the learning of new skills, is reduced. For example, during the building of the IDDA end-system:

- the word-processing of both the questionnaires and the specialist help could be allocated to the unit's secretarial support staff, who already have the necessary typing skills,
- the interaction with the computer-based tools to build the required IDDA end-system could be one of the tasks for the trainee specialist, who knows the domain sufficiently well to recognise and correct any errors,
- whereas, the designing of the actual questionnaires and the specification of any help could be left to the expert, who already devises such questionnaires for current manual studies.

As the experts are aware of the investigations they wish to undertake as well as being knowledgeable over what data and information needs to be collected to enable the various analyses to be carried out, they are ideally placed to design the questionnaires. They also know the daily working practices of the unit, the normal procedures that are undertaken and their sequencing, the terminology of the domain, and the likely associations or influences that certain characteristics or tests have on each other. Furthermore, they know the current research interests within their domain. Thus, they can plan and instigate studies to complement or query such theories as well as undertaking investigations which explore new facets of the field.

Therefore, by basing the IDDA system on the questionnaires designed by the domain expert, it can be integrated smoothly into the daily routine of the unit with little, if any, delay in the data gathering and with the minimal amount of reorganisation of activities. This similarity with the manual procedures and the daily routine also gives users confidence and reduces the anxiety created when they are confronted with something new and unknown. In a similar fashion, the use of standard questionnaires provides users with extra support, as they know what information or action to expect next (Rector, 1989).

Consequently, the decision was to construct each IDDA end-system from the questionnaires designed by a specialist. It provides facilities to store collected data in an orderly fashion whilst reducing the likelihood of invalid data entries and incomplete records. In addition, it enables quick reviews of all the stored information. It also permits statistical analyses to be undertaken on the collected data by providing in-built statistical facilities. Thus, some analyses can now be undertaken which had in the past been too tedious or complex to carry out by hand. Furthermore, by computerising the analysis stage, the speed and accuracy with which results can be obtained is increased dramatically (see Chapter 5).

The actual running of the IDDA end-system follows a similar approach to the building process in the allocation of duties:

- data entry from the paper records can be performed by the support staff,
- any general enquiries and reviews can be carried out by the trainee specialist,
- the detailed analyses and investigations can be undertaken by the experts.

Thus, the relative strengths of the individuals within the unit can be used to best effect. This would therefore save time, reduce errors, increase the users' acceptance in the end-system as well as their willingness to be involved in the system's construction. It would also encourage the feeling of participating within a team and of ownership of the end-system throughout the whole unit, since everyone has a part to play in both the development and the continuing daily operation of the IDDA end-system.

With regard to the selection of the interaction style to be adopted for the tools and the IDDA system, the decision was to design and use techniques which would be as intuitive as possible to users. This was to ensure that the demands placed on users would be minimal and that both systems would be very easy to operate and quick to learn, with the result that they would be operated more efficiently and effectively, even by users with little or no computer experience (see Chapter 3 and Chapter 4).

Consequently, the decision was made that the mode of data entry would be based solely on a keyboard. This allows for the entry of both free-text information as well as for the selection of options without the necessity of switching between different devices. It also enables the utilisation of any typing experience that the users might possess.

The dialogue styles, i.e. menus and prompt questions, were also chosen to match the requirements of the tasks and to aid users during operation, for example, by reducing the memory load and by increasing the familiarity with the specialist questionnaires. Extensive error checking of the information entered by a user was included as it gives a user extra confidence and reduces the number of errors in the stored data. In addition, there is the ability for an IDDA system to have user-defined help linked to the questions in the assessment questionnaires, thereby enabling experts to explain specialist terminology and techniques in their own words.

With regard to the analysis facilities, an IDDA system contains the main characteristics that Akazawa et al (1991) reported that their clinicians had requested:

- it uses an interactive mode that is based on menu and question-and-answer style interfaces,
- there is Help available at the request of users. Moreover, not only is it available, it is displayed on the second monitor so as not to cause any interference to the main screen,
- users can select from a wide range of statistical tests,
- users can add, amend or review data stored in the assessment databases.

Therefore, through these facilities that are offered by an IDDA system, it is anticipated that physicians will carry out their own analyses and use statisticians only to check the methods selected, if confirmation of the results is needed.

It must, however, be clearly emphasised that any decisions or conclusions drawn from the analysis of information are the responsibility of the expert. An IDDA end-system does not dictate to specialists but merely provides them with facilities to investigate their field, thus enabling more informed and hence better decisions to be made. Therefore rather than trying to mimic human specialists as an expert system might attempt, an IDDA end-system is just another aid to use as and when required.

Consequently, an IDDA end-system works along-side the expert and should therefore not cause any adverse effect on a user's behaviour in either the short or longer term. It neither usurps nor threatens a user's position or status. In this manner, an IDDA end-system will not have to address the important legal, social or ethical issues that have been raised during the implementation of various expert and decision-making systems. It can therefore be perceived as merely being a tool to assist specialists in their work.

7.4.1 Future developments and further research

There are numerous avenues which could be explored either by using the tools and IDDA systems they build or by studying these two systems themselves. Many of these interesting issues have already been highlighted throughout this thesis, especially in Chapter 1 and Chapter 6. For example, the effectiveness of the systems and their impact on users and specialist domains need to be analysed after a substantial period of operation in a number of different working environments and specialisms.

The investigations can be undertaken at a low level, i.e. by examining whether there are any similarities between the new facilities requested by various users, or by determining whether the working practices and daily routines have been changed. Alternatively, they could be at a higher level, for example, reviewing whether explicit management strategies or decision pathways have been established and whether these have led to the specialist fields becoming more formalised and standardised. In addition, an analysis could be carried out to determine whether there have been any improvements in either the health care of patients due to the introduction of these systems or else in the acceptability and credibility of computer systems within medical fields.

Other studies could review the transportability of the IDDA system and whether the IDDA system increases the likelihood of carrying out multi-site investigations or of 'pooling' data collected from various groups. However, the likely direction that would capture the most interest at present would involve the current research being undertaken into automatic knowledge discovery and data 'mining' techniques and the subsequent testing or application of these methods to the data collected within the various IDDA databases.

Consequently, it is hoped that the methodology proposed within this project will encourage and support further developments and research not only within the computer community but also by investigators within medicine.

7.5 The advantages of the methodology

IDDA end-systems provide domain experts with facilities to explore and investigate their fields. It uses questionnaires devised by an expert to collect the required data in an orderly and standardised manner. Questionnaires, whether paper-based or computerised, have been shown to be beneficial as they ensure that a thorough case-history is recorded. However, unlike a paper-based approach, computerised questionnaires can check the validity of the values entered, can insist on particular questions always being answered, and can quickly provide relevant domain specific help for whoever may require further explanations of procedures or terminology.

Moreover, with the inbuilt statistical capabilities of the IDDA system, an expert has the ability to select and undertake reviews of the data stored, whilst remaining within the same system. Hence, as analyses, especially complex ones, can be undertaken faster and in a more convenient manner than is the case currently, experts will be encouraged to carry out further investigations. Therefore, associations and influences between data items, which had previously been unknown or hard to appreciate, are more likely to be uncovered.

However, just like any other tool, the results obtained from any analyses are dependent upon the skill of users to select the relevant data and the technique to use. This skill is not reliant upon computer or computing knowledge but on the ability of an expert to manipulate the stored information to prove or disprove a theory or to examine an unusual trend. Therefore users concern themselves with the specialist problem at hand rather than trying to understand the technical workings of the computer system. The IDDA provides the facilities to carry out these analyses quickly, easily and accurately, but the formation of the analyses and the justification of the conclusions are left to domain specialists, who has the ability to include those influences that are outside the scope or remit of computer systems.

One advantage of computerising the methodology through building a suite of appropriate computer-based tools is to enable domain experts to construct their own IDDA end-system, thus permitting experts to take full control of their studies. Consequently, experts will no longer be reliant on the availability of computer scientists.

Furthermore, as domain experts determine the details of an investigation to be undertaken, i.e. the data to be gathered and the methods to be used, and as the computer system is then built from these specifications, the resultant system will thus be appropriate for an expert's intended study. Also the speed of constructing this system is within the control of domain experts since they provide the tools with the information from which an IDDA end-system is constructed. Therefore, once a study is devised, a system can be produced as and when it is needed by the specialist. Moreover, as one of its tasks is to gather and store the data for the investigation in an orderly fashion, an IDDA end-system can assist specialists in their work as soon as a study is initiated, rather than having the normal lead-in time before a new computer system can undertake any productive work.

A further advantage of experts designing and building their own IDDA end-system is that they can ensure that, with the introduction of the computer system, there is minimal disruption to the daily working practices of the specialist unit and, in medical fields, to the patient-physician relationship. They can not only dictate when a system is to be used, i.e. during a patient session or afterwards, but they will also have specified the procedures to be used to acquire data and the order in which these procedures are to be carried out, thus ensuring that the computerised procedures conform to the facilities available and the established work patterns of the unit. Consequently, IDDA end-systems can be constructed which are appropriate to the local environment.

In addition, this ability allows studies to be devised which involve a number of dispersed centres. With an IDDA end-system being based on data and facts rather than on knowledge, the differing opinions, preferences and beliefs of individual experts are unlikely to present many difficulties during a system's construction. Each centre can develop its own system which fits into the local group's working practices and the terminology that is normally used, but which still gathers all the necessary data for a study. The results obtained at one centre can then be confirmed or questioned by the results gained in another. Therefore this approach allows more detailed investigations to be undertaken on a more widespread scale.

Moreover, if appropriate, the case-histories from all of the centres involved in a study can be combined to produce one large data set. This would obviously result in such a data set becoming available much more quickly than any of the individual sites could possibly achieve alone and would be beneficial for those analyses which require large data sets or for increasing confidence in the results obtained. It must be stressed that this 'pooling' of data would only be appropriate under particular conditions otherwise the statistical validity of the analyses would be questionable. However, in certain situations, it could be appropriate and the benefits could be realised.

It may also be necessary to gather data from various outlying centres. As an IDDA system can be loaded and run on a machine that is easily carried, for example, a portable PC, this requirement can be met. The analysis or review of data can be undertaken back at the main centre once the new data has been appended to the other case-histories. This ability of system transportability would be especially useful to medical experts who may have the responsibility for vast areas or for dispersed populations, for example in Papua New Guinea or India.

With regards to upgradability, the structure of the IDDA end-system is such that other methods of analysing or reviewing data can be easily integrated. The interaction style used for these new analyses would follow the same techniques as currently used in the IDDA system. Therefore, the inclusion of any additional facilities will be relatively seamless to the users. In fact, it would be possible to allow experts to request just those facilities they require for their specialist field, thereby enabling the customisation of an IDDA end-system and preventing 'over-loading' the users with facilities they would not use.

The evaluations undertaken in Chapter 6 reported that the evaluators perceived the IDDA end-system as not only being very useful, especially with its ability to provide in-built statistical facilities, but that such a system would encourage further investigation and research to be undertaken and, thus would aid in improving health care. They also believed that the ability for specialists to construct their own system was extremely desirable and highly beneficial. They felt that a specialist would then be in control of the whole investigation, from the initiation through the implementation to the operation and the analysis of data and finally, to the interpretation of results and the drawing of conclusions or recommendations. They also liked the power this approach gave researchers to instigate new studies with appropriate computer support whenever required, without delay and with the minimal amount of cost. Hence they believed that more studies would be carried out. These would be run more quickly, investigated more accurately, thoroughly and in more depth with the result that new insights and knowledge of the specialist domain would be uncovered.

It was therefore anticipated that such studies will ultimately lead to the emergence of more formalised and standardised fields. However, in the shorter term, it was appreciated that the investigations are more likely to concentrate on the current practices and procedures undertaken in the specialist domain, to record and reveal the various strengths and weaknesses of each. This should result in a reduction in the running costs as irrelevant tests are eliminated or similar tests rationalised. In a medical field, such an effect is not a simple monetary saving, but would lead to improvements in health care, i.e. a decrease in the amount of stress, discomfort and pain suffered by patients undergoing treatment, whilst also enabling more patients to be treated due to the savings of both time and money.

7.6 Advantages computers can bring to medicine

Most domains in which humans operate are inherently dynamic in that they are constantly evolving and changing. This is especially true of fields such as medicine. It is out of the question to expect any single person to read, let alone absorb, the details of all of the medical research projects currently being undertaken. This continual renewal and expansion of information and knowledge has resulted in specialisms forming within each field and has forced physicians to become more dependent on advice from other sources when they are presented with problems outside their own area of expertise (Shortliffe et al, 1984). However, as previously explained in Chapter 6, even these sub-fields are now growing too rapidly and further division would be undesirable in terms of economics, practicality and the ability to generalise back from the specialised view point to take into account more global issues. These specialists also require extensive training and are, in fact, no solution for poor or large, sparsely populated countries, as Hand (1985) points out, 'what use is the expert if he is 2000 miles away and the only means of transport is a mule?'. Therefore workers within such fields require assistance, especially those who make decisions and those who carry out research.

The information processing capabilities of computers not only enable the rapid manipulation of large quantities of data in an error-free way, but also encourages the widespread dissemination of medical knowledge. The ability of physicians to access and question large stores of data enhances the possibility of determining more effective and efficient processes for managing patients. In fact, as Fox et al (1990) state, 'there is a growing belief that information and decision support techniques will be crucial to future improvements in standards of patient care'.

However, the main objective should be to create a human-machine team which performs better than any individual team member alone. Consequently, activities should be identified for each member in which each has a relative advantage over the other. In this manner, a computer should not be involved in tasks in which its contributions are insignificant or questionable or where the interaction required places too heavy a demand on the user. Hence the design of a computer system should be driven by the intended user group, the task and the environment and be guided by what a computer should do rather than what it can do.

This is particularly important in medicine. As more is understood of the complex and changing nature of medical knowledge, it is clear that a co-operative relationship must exist between physicians and computer-based decision tools. This approach of designing decision support systems which work with human experts seems likely to lead to much greater success than has been achieved in the past. Such co-operative systems can, to some extent, act as de-biasing mechanisms by providing access to large databases of factual information which could be searched in a comprehensive and unbiased manner.

Tasks, such as information acquisition, management and processing, are ideally suited to a computer system. Hence an appropriate, well-designed system will reduce the time and effort required to undertake an investigation by reducing the information processing load on the decision-maker (Todd and Benbasat, 1992).A user can then decide whether to invest these savings either:

a) in finding better quality decisions and solutions through greater information use and more complete analysis,
b) in solving other problems, including those activities not supported by the computer system.

However, to achieve either of the above, the effort required to interact with the computer system must be less than the effort to undertake the task unaided and the decision-maker must be dissatisfied with the current level of decision quality to be motivated enough to contemplate an alternative.

Taylor et al (1971) discovered that humans tend to be conservative in their processing of information and that some clinicians seem to require much more information than others before they make a decision. In Fisher (1994), Hart agreed commenting that 'Uninformed activity is dangerous' and he talked of the clinician's dilemma of having to be right. 'It is why junior doctors ask for more information than senior ones. Uncertainty is expensive'. In certain situations, tests are merely being carried out to confirm results already obtained from previous tests. The result is the increasing cost of health care, for example, currently 13.2% of America's national product goes on health care provision. Therefore Fisher (1994) believes that these escalating medical costs throughout the world need to be checked by the use of technology.

Computers are immune to the boredom, fatigue, and situational and interpersonal distractions which detract from human performance. Consequently, their output and responses will be consistent and reliable no matter what task they are requested to carry out, when or how many times. Therefore, if computer systems designers could provide information processing support, decision-makers could be left to decide how to make best use of the capabilities of a system. There is evidence that such assistance and support creates a 'climate' in which doctors are stimulated and motivated towards doing the work correctly, leading to improvements in decision-making (Adam et al, 1986). For example, de Dombal (1984) discovered that doctors, possibly stimulated by the novel experience of a computer working alongside them, possibly deriving benefit from the feedback, or possibly simply benefiting from the disciplined data collection necessary to operate such computer systems, often improve their own clinical performance. De Dombal (1984) calculated from UK trials in 1984 that this improvement in doctors' performance could lead to a freeing of resources worth £50,000 - £100,000 per hospital per year for other uses.

However, it is as serious a mistake to exaggerate the usefulness of computers as it is to under-estimate the value of the computer systems. They are highly unlikely to ever replace a human decision-maker, for example, in being able to deal with unexpected situations or to deal with social and ethical issues that are often so important in 'real-world' medical decision-making. Hence, computer systems should be designed to take full advantage of their strengths, thereby releasing human decision-makers to concentrate on those aspects of problem-solving in which humans excel.

Consequently the aim should be to provide a combined system which produces far better decisions than either a human or a computer could have achieved on their own.

Currently domain experts, who are naive computer users, do not have the tools available to them to construct their own computer systems with which to investigate their specialist fields. Moreover, computer systems which are constructed for specialists often do not take into account all the environmental factors, the task or the user requirements, constraints and issues that are present within the domain. For example, requiring that users change their daily working practices and use unfamiliar terminology, a tendency to provide only limited facilities for reviewing the stored information, and the most common, a lack of any inbuilt statistical analysis capabilities which are essential when undertaking empirical investigations. This diminishes the usefulness and acceptability of such a computer system to the intended users. Therefore there has been an inclination to use manual methods to carry out investigations and the required analyses. This obviously restricts the number of studies that are undertaken and the detail, accuracy and depth of the analyses. It also removes tasks from the computer system for which it is ideally suited, e.g.that of data storage and data manipulation.

An appropriate computer system could assist such investigations immensely. By concentrating on data and facts rather than on experts' knowledge and experience, many of the difficulties that have confronted expert and knowledge-based systems can be avoided. With non-formalised fields, this movement away from individual personal beliefs, biases, and experiences is essential for the production of any computer system that is to be of widespread use. Therefore there is a need to be able to build computer systems which can undertake the required empirical studies.

Furthermore, as one investigation initiates another, there is a need for the ability to quickly and easily construct these systems as and when they are required. To provide this flexibility and to ensure that an end-system can integrate into a working environment, the initiators of the investigations should also be the builders of the end-systems. However they are likely to be naive computer users and therefore the processes of developing and constructing a system as well as operating an end-system, all have to be within the capabilities of a user with little or no prior computing experience.

This research has proposed a methodology which provides such a suite of computer-based tools to enable specialists to take full control of developing their own IDDA end-systems for their investigations and any subsequent studies that they may wish to undertake. Hence more investigations are likely to be carried out and, with the inbuilt statistical capabilities of the IDDA system, the data collected can be reviewed and analysed in more detail and depth, thereby resulting in further information, associations and knowledge of the specialist field being revealed. Thus, in addition to users immediately acquiring assistance in the collection and retrieval of data with an IDDA end-system, they also have the ability to investigate and explore, in detail, their specialist domain, their decision-making paths and their problem-solving processes. Consequently, by utilising the tools and the IDDA end-system they produce, users would have the ability to actively participate in researching and advancing their field and in evolving a more formalised and standardised specialist domain.

Accordingly, as Lenat and Feigenbaum (1991) have pleaded, 'build some intelligent interfaces that allow us to write programs more easily, or synthesise ideas more rapidly etc. Then let these improved man-machine systems loose on the problem of achieving AI, whichever goal we choose to define it. In other words, instead of tackling the AI task right away, let's spend time getting prostheses that let us be smarter, then we'll come back to working on "real" AI.'

References

Adams, I.D., Chan, M., Clifford, P.C., Cooke, W.M., Dallos, V., De Dombal, F.T., Edwards, M.H., Hancock, D.M., Hewett, D.J., McIntyre, N., Somerville, P.G., Spiegelhalter, D.J., Wellword, J. and Wilson, D.H., 1986, 'Computer aided diagnosis of acute abdominal pain: a multicentre study', Brit. Med. J., Vol.293, p800-804

Addis T.R., 1985, Designing knowledge based systems, London: Kogan Page

Addis, T.R., 1987, 'A framework for knowledge elicitation', Proc. of the First European Workshop on Knowledge Acquisition for Knowledge-based systems, Reading Uni. p1-7

Aitkenhead, A.M. and Slack, J.M., 1990, Issues in cognitive modelling, London: Lawrence Erlbaum, Assoc. Ltd.

Aitkinson, R.C. and Shiffrin, R.M, 1971, 'The control of short-term memory', Scientific American, Vol.224, p82-90

Aitkinson, R.C. and Shiffrin, R.M., 1977, 'Human memory: A proposed system and its control processes', Human memory: Basic processes, (ed) Bower, G.H., New York: Academic Press, p1-115

Aitkinson, R.L. Aitkinson, R.C. and Hilgard, E.R., 1983, Introduction to psychology, New York: Harcourt Brace Jovanovich

Akazawa, K., Shimada, M., Moriguchi, S., Fujisawa, K., Odaka, T. and Nose, Y., 1991, 'Interactive statistical analysis system for clinical investigators', Med. Inform., Vol.16, No.4, p347-354

Akman, V. and Ten Hagen, P.J.W, 1989, 'The power of physical representations', AI Magazine, Fall, p49-65

Alonso, F., Mate, J.L. and Pazos, J., 1990, 'Knowledge engineering versus software engineering', Data and Knowledge Engineering, Vol.5, p79-91

Altman, D.G., 1991, Practical statistics for medical research, London: Chapman and Hall

Amalberti, R., Carbonell, N. and Falzon, P., 1993, 'User representations of computer systems in human-computer speech interaction', Int. J. Man-machine Studies, Vol. 38, p547-567

Anbar, M., 1987, 'Artificial Intelligence and its impact on medicine - can a computer replace the doctor?', Computers in Medicine, Maryland, USA: Comp. Sci. Press

Anderson, J.R., 1982, 'Acquisition of a cognitive skill', Psychology Review, Vol.89, No.4, p369-406

Andriole, S. J., 1985, 'Applied Artificial Intelligence in perspective', Applications in Artificial Intelligence, (ed) Andriole, S.J., Princeton, N.J.: Petrocelli Books

Baddeley, A.D., 1990, 'Domains of Recollection', Issuses in Cognitive Modelling, (ed) Aitkenhead, A.M. and Slack, J.M., London: Lawrence Erlbaum Associates Ltd., p209-229

Bartlett, F.C., 1958, Thinking, New York: Basic Books

BCS Committee of the Specialist Group on Expert Systems Interim Report, 1983

Ben-Bassat, M., 1985, 'The role of expert systems in clinical diagnosis', Applications in Artificial Intelligence, (ed) S.J. Andriole, Princeton, N.J.: Petrocelli Books, Inc.

Berg-Cross, G. and Price, M.E., 1989, 'Acquiring and managing knowledge using a conceptual structures approach: introduction and famework', IEEE Trans. on Systems, Man and Cybernetics, Vol.19, No.3, p513-27

Bergson, H., 1911, Evolution creatice, translated Mitchell, A., London: Holt

Berry, D.C., 1987, 'The problem of implicit knowledge', Expert Systems, Vol.4, No.3, p144-151

Berry, D.C., 1994, 'Involving users in expert system development', Expert Systems, Vol. 11, No.1, p23-28

Berry, D.C. and Broadbent, D.E., 1987, 'Expert systems and the man-machine interface. Part two: The user interface', Expert System, Vol.4, No.1, p18-28

Berry, D. and Hart, A., 1990, 'Evaulating expert systems', Expert Systems, Vol.7, No.4, p199-208

Berry, D. and Hart, A., 1991, 'User interface standards for expert systems: are they appropriate ?', Expert Systems with applications, Vol.2, p245-250

Bielawski, L. and Lewand, R., 1991, Intelligent systems design, New York: John Wiley and Sons

Birmingham, W. and Klinker, G., 1993, 'Knowledge-acquisition tools with explicit problem-solving models', The Knowledge Engineering Review, Vol. 8, No. 1, p 5-25

Bjerragaard, B., Brynite, S., Holtz-Christensen, J., Kalaja, E., Lund-Kristensen, J., Hilden, J., de Dombal, F.T. and Horrocks, J.C., 1976, 'Computer aided diagnosis of acute abdomenal pain: a system from Leeds used on Copenhagen patients', Decision-making and Medical Care, (eds) de Dombal, F.T. and Gremy, F., North-Holland: Elsevier Science

Blascovich, J., 1987, ' The clerk who never forgets - database management', Computers in Medicine, (ed) Anbar, M., Maryland, USA: Comp. Sci. Press

Boehm-Davis, D.A. and Ross, L., 1992, 'Program design methodologies and the software development process', Int. J. Man-Machine Studies, Vol.36, p1-19

Boose, J.H., 1985, 'A knowledge-acquisition program for expert systems based in Personal Construct psychology', Int. J. Man-Machine Studies, Vol.23, p495-525

Booth, P.A., 1991, 'Errors and theory in human computer interaction', Acta. Psychologica, Vol.78, p69-96

Borenstein, N.S., 1985, The design and evaluation of on-line help systems, Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA.

Bradbury, A.R., 1991, 'Computerised medical records: defining a standard without the computer', Med. Inform. Vol.16, No.3, p279-286

Braunstein, B., Lauer, T.W. and Doanne, D.P., 1991, 'Information requirements prototyping for decision support systems', J. Infor. Tech., Vol.6, p26-33

Brender, J. and McNair, P., 1989, 'Watch the system. An opinion on User Validation of computer-based decision support systems in medicine', submitted to MEDINFO-89

Bridge, R.E. and Williamson, G.L., 1981, 'Complete computer systems for general practioners', BSC '81 Information Technology for the eighties, (ed) Parslow, R.D.,p200-206

Brooking, A.G., 1986, 'The analysis phase in development of knowledge-based systems', Artificial Intelligence and Statistics, (ed) Gale, W.A., AT and T Bell Labs.

Brooks, R., 1991, 'Intelligence without reason', 12th International Joint Conference on Artificial Intelligence, (eds) Mylopoulos, J. and Reiter, R., San Mateo, Calif.: Morgan Kaufmann

Buchanan, B.G. and Shortliffe, E.H., 1984, Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project, New York: Addison-Wesley

Burgess, C.G., 1991, 'A graphical database-querying interface for casual, naive computer users, Int. J. Man-Machine Studies, Vol.34, p23-47

Byrd, T.A., Cossick, K.L. and Zmud, R.W., 1992, 'A synthesis of research on requirements analysis and knowledge acquisition techniques', MIS Quarterly, Vol.16, No.1, p.117-138

Card, S., Moran, T.P. and Newell, A., 1983, The psychology of human-computer interaction, Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Carey, J.M., 1990, 'Prototyping: alternative systems devlopment methodology', Info. Software Tech., Vol.32, No.2, p119-126

Carroll, J.M. and Carrithers, C., 1984, 'Training wheels in a user interface', Communications of the ACM, Vol.27, p800-806

Carroll, J.M., and Rosson, M.B., 1985, 'Usability specifications as a tool in iterative design', Advances in human computer interaction, (ed) Hartson, H.R., Ablex Pub. Cor., p1-29

Carroll, J.M. and Campbell, R.L., 1989, 'Artifacts as psychological theories: the case of human-computer interaction', Behaviour and Information Technology, Vol.8, No.4, p247- 256

Chalmers, T.C., Frank, C.S. and Reitmen, D., 1990, 'Minimising the three stages of publication bias', J. of American Medical Association, Vol.263, p1392-1395

Chapinis, A. and Budurka, W.J., 1990, 'Specifying human-computer interface requirements', Behaviour and Info. Tech., Vol.9, No.6, p479-492

Cherry, J.M., Fischer, M.J., Fryer, B.M. and Steckham, M.J., 1989, 'Modes of presention for on-line help: full-screen, split-screen and windowed formats', Behaviour and Info. Tech., Vol. 8, No. 6, p405-416

Clare, A, 1976, Psychiatry in dissent, London: Tavistock Pub.

Clegg, C., Warr, P., Green, T., Monk, A., Kemp, N., Allison, G. and Lansdale, M., 1988, People and Technology: How to evaluate your company's new technology, Chichester: Ellis Horwood

Coates, R.B. and Vlaeminke, I., 1987, Man-Computer interfaces: An introduction to software design and implementation, London: Blackwell Scientific Publ.

Cohen, P.R., 1987, 'Numeric and symbolic reasoning in expert systems', Advances in Artificial Intelligence II, (ed) Du Boulay, B., Hogg, D. and Steels L., North-Holland: Elsevier Science

Coles, E.C., 1973, A guide to Medical Computing, London: Butterworths

Collste, G., 1992,'Expert systems in medicine and moral responsibilities', J. Systems and Software, Vol.17, No.1, p15-22

Collins, H.M., Green, R.H. and Draper, R.C., 1985, 'Where's the expert in expertise?: Expert systems as a medium of knowledge transfer', Expert systems '85 Brit. Comp. Soc. Workshop Series, (ed) Merry, M., New York: Cambridge Uni. Press, p323-334

Conklin, J., 1988, 'Hypertext: an introduction and survey', Computer-supported cooperative work: a book of readings, (ed) Grief, I., San Mateo, Calif.: Morgan Kaufmann, p423-475

Cordingley, E.S., 1989, 'Knowledge elicitation techniques for knowledge based systems', Knowledge elicitation, (ed) Diaper, D., Chichester: Ellis Horwood., p 89-177

Cuesters, W., De Moore, G., Bonneu, R. and Schilders, L., 1992, 'Training of health care personnel towards the implementation and use of electronic health care records using integrated imaging technology', Med. Inform., Vol.17, No.5, p215-223

Cushing, M., 1991, 'The Clintrac Medical Data System', M.D. Computing, Vol.8, No.5, p326-328

Davies, M. and Hakiel, S., 1988, 'Knowledge harvesting: A practical guide to interviewing', Expert systems, Vol.5, No.1, p42-49

Davis, R. (ed), 1989a, 'Part one: Expert Systems: How far can they go ?', AI Magazine, Spring, p61-67

Davis, R. (ed), 1989b, 'Part two: Expert Systems: How far can they go ?', AI Magazine, Summer, p65-77

Davis S.A. and Bostrom, R.P., 1993, 'Training end-users: An experimental inmvestigation of the roles of the computer inteface and training methods', MIS Quarterly, Vol.17, No.1, p1-8

de Dombal, F.T., Leaper, D.J., Stainland, J.R., McCann, A.P. and Horrocks, J.C., 1972, 'Computer-aided diagnosis of abdominal pain', Brit. Med. J., Vol.2, p9-13

de Dombal, F.T., 1978, 'Medical diagnosis from a clinician's view point', Meth. of Infor. in Med., Vol.17, p28-35

de Dombal, F.T., 1983, 'Towards a more objective evaluation of computer-aided decision support systems', MEDINFO-83, (eds) Van Bemmel, Ball and Wiegertz, North-Holland: Elsevier Science, p436-439

de Dombal, F.T., 1984, 'Clinical decision-making and the computer: consultant, expert or just another test?' Brit. J. Healthcare Computing, Vol.1, No.1, p7-14

de Dombal, F.T., 1988, Data-based aids (video), D321 Professional Judgment, The Open University

de Dombal, F.T., Dallos, V. and McAdam, W.A., 1991, 'Can computer aided teaching packages improve clinical care in patients with acute abdominal pain ?', Brit. Med. J., Vol.302, p1495-7

Diaper, D., 1989, 'Designing expert systems: from Dam to Beersheba', Knowledge elicitation principles, techniques and applications., (ed) Diaper, D., Chichester: Ellis Horwood, p15-46

Dodson, D.W. and Shields, N.E., 1978, 'Development of user guidelines for ECAS display design', Vol.1, report No. NASA-CR-150877, Essex Corp. Huntsville

Donabedian, A., 1980, Explorations in Quality Assessment Monitoring, Health Administration Press, Ann Arbor, Michigan

Douglas, S.A., 1982, 'Learning to text edit: semantics in procedural skill acquisition', Doctoral Disseration, Standford Uni.: Palo Alto, C.A.

Du Bois, P., Hermanus, J.P. and Cantraine, F., 1989, 'A system for computer-aided diagnosis', Med. Inform., Vol.14, No.1, p13-23

Draper, S., 1984, The nature of expertise in UNIX, (Technical Report, No. 8402), San Diego: University of California, Institute for Cognitive Science

Duchessi, P. and O'Keefe, R.M., 1992, 'Contrasting successful and unsuccessful expert systems', European J. of Operational Research, Vol.61, p122-134

Duda, R.O. and Shortliffe, E.H., 1983, 'Expert systems research', Science, Vol.220, p261- 268

Duncker, K., 1926, 'A Qualitative (Experimental and Theoretical) study of productive thinking (solving of comprehensible problems)', J. Genetic Psychology, Vol.33, p642- 708

Edmonds, E., 1990, 'Human-computer interface evaluation: not user-friendliness but design for operation', Med. Inform., Vol.15, No.3, p253-260

Einhorn, H.J. and Hogarth, R.M., 1981, 'Behavioral decision theory: processes of judgement and choice', A. Rev. Psychol., Vol.32, p53-88,

Elstein, A.S., Sulman, L.S. and Sprafka, S.A., 1978, Medical problem solving: an analysis of clinical reasoning, Cambridge: Harvard Uni. Press

Essens, P.J.M.D., McCann, C.A. and Hartevelt, M.A., 1991, 'An experimental study of the interpretation of logical operators in database querying', Acta Psychologica, Vol.78, p201- 225

Evans, J.B.St., 1982, The psychology of deductive reasoning, London: Routledge and Kegan Publ.

Evans, J.St.B.T., 1983 'Selective processes in reasoning', Thinking and reasoning: Psychological approaches, (ed) Evans, J.St.B.T., London: Routledge and Kegan Publ., p135-164

Evans J. St. B. T., 1987a, 'Human biases and computer decision-making: a discussion of Jacob et al', Behaviour and Information Technology, Vol.6, No.4, p483-487

Evans J. St. B. T., 1987b, 'Beliefs and expectations as causes of judgemental bias', Judgement Forecasting (ed) Wright, G. and Ayton, P., New York: John Wiley and Sons, p31-47

Evans, J.St.B.T, 1989, 'Problem solving, Reasoning and Decision-making', Cognitive Psychology: European Perspectives Vol. 1, (ed) Baddeley, A. and Bernsen, N.O., London: Lawrence Erlbaum Ltd., p85-103

Feigenbaum E.A. and McCorduck P., 1983, The Fifth Generation, New York: Addison-Wesley

Fieschi, M., 1990, Artificial Intelligence in Medicine, London: Chapman and Hall

Fischer, G. and Leuke, A.C., 1987-88, 'Construction kits and design environments: steps towards human problem-domain communications', Human-computer interaction, Vol.3, p179-222

Fischhoff, B., Slovic, P. and Lichtenstein, S., 1977, 'Knowing with certainty: The appropriateness of extreme confidence', J. Exp. Psychol. Hum. Percept. Perform., Vol.3, p552-564,

Fisher, J., 1985, 'Medical information and patient power', Computer applications in medical care, (ed) Ackerman, M.J., Los Angeles, Calif.: IEEE Comp. Soc. Press, p655-661

Fisher, P., 1994, 'Windows on more than the soul', The Guardian, Feb. 17th, p19

Flagle, C.D., 1982, 'Chairman's introduction - Some terminology of evaluation', Proc. 6th Ann. Symp. Comp. Appl. in Medical Care, Comp. Soc. Press, IEEE Comp. Soc, Los Angeles: Calif., p439-440

Fowler, C., Macaulay, L. and Siripoksup, S., 1987, 'An evaluation of the effectiveness of the adaptive interface moudule (AIM) in matching dialogues to users', People and Computers III (ed) Diaper, D. and Winder, R., New York: Cambridge Uni. Press, p345-361

Fox, J., 1984, 'Formal and knowledge-based methods in decision technology', Acta. Psychologica, Vol.56, p303-331

Fox, J., Barber, D.C. and Bardhan, K.D., 1980, 'Alternatives to Bayes: A quantitative comparison with rule-based diagnosis', Methods of Information in Medicine, Vol.19, No.4, p210-215

Fox, J., Glowinski, A., Gordon, C., Hajnal, S. and O'Neil, M., 1990, 'Logic engineering for knowledge engineering: design and implementation of the Oxford System of Medicine', Artificial Intelligence in Medicine, Vol.2, p323-339

Fox, J. and Krause, P., 1992, 'Qualitative frameworks for decision support: lessons from medicine', The Knowledge Engineer Review, Vol.7, No.1, p19-33

Friedman, R.B. and Gustafson, D.H., 1977, 'Computers in clinical medicine: a critical review', Comput. Biomed. Res., Vol.10, p199-204

Frohlich, D.M., Crossfield, L.P. and Gilbert, G.N., 1985, 'Requirements for an intelligent form-filling interface', People and Computer: Designing the Interface, (ed) Johnson, P. et al, New York: Cambridge Uni. Press, p102-117

Gage, J.S., 1992a, 'Mathematics and judgement: a review of the computer simulation program GAS man', M.D. Computing, Vol.9, No.1, p54-56

Gage, J.S., 1992b, 'Uncertainty and critical incidents', M.D. Computing, Vol.9, No.6, p355-357

Gage, J.S., 1993a, 'Notes from Stony Brook', M.D. Computing, Vol.10, No.6, p370-372

Gage, J.S., 1993b, 'Subjective probability', MD Computing,Vol. 10, No. 5, p297-298

Gaines, B.R., 1987, 'An overview of knowledge-acquisition and transfer', Int. J. Man-Machine Studies, Vol.26, p453-472

Gaines, B., 1988, 'An overview of knowledge acquisition and transfer', Knowledge acquisition for knowledge based systems, (ed) Gaines, B. and Boose, J., London: Academic, p3-23

Gammack, J.G., and Young, R.M., 1984, 'Psychological techniques for eliciting expert knowledge', Research and Development of expert systems, (ed) Bramer, M.A., Cambridge, UK: Cambridge Uni. Press

Garg-Janaradan, C. and Salvendy, G., 1987, 'A conceptual framework for knowledge elicitation', Int. J. Man-Machine Studies, Vol.26, p521-531

Garg-Janardan, C. and Salvendy, G., 1988a, 'The contribution of cognitive engineering to the design and use of expert systems', Behaviour and Information Technology, Vol.7, No.5, p323-342

Garg-Janardan, C. and Salvendy, G., 1988b, 'A structured knowledge elicitation methodology for building expert systems', Int. J. Man-Machine Studies, Vol.29, p377-407

Gerlach, J. H. and Kuo, F., 1991, 'Understanding Human-computer interaction for information systems design', MIS Quarterly, Vol. 15, No. 4, p 527-548

Gierl, L. and Stengel-Rutkowski, S., 1994, 'Integrating consultation and semi-automatic knowledge acquisition in a prototype-based architecture: Experiences with dysmorphic syndromes.', Artificial Intelligence in Medicine, Vol. 6, No. 1, p29-49

Gordon, S.E., Gill, R.T. and Dingus, T.A., 1987, 'Designing for the user: The role of human factors in expert system development.', AI Appli. in Nat. Resource Man., Vol.1, No.1, p35-46

Gordon, S.E., 1989, 'Theory and methods for knowledge acquisition', AI Applications, Vol.3, No.3, p19-30

Gore, S.M. and Altman, D.G., 1990, Statistics in Practice, London: British Medical Association

Gould, J.D. and Lewis, C.H., 1985, 'Designing for usability and what designers think', Communications of the ACM, Vol.28, p300-311

Greenberg, S. and Witten, I., 1983, 'Adaptive personalised interfaces - A question of viability', Behaviour and Information Technology, Vol.4, p31-35

Gremy, F., 1989, 'Human meaning of medical informatics: reflections on its future and trends', Med. Inform., Vol.14, No.1, p1-11

Grover, M.D., 1983, 'A pragmatic knowledge acquisition methodology', Int. J. Conf. on AI, p436-438

Gruber, T.R., 1988, 'Acquiring strategic knowledge from experts', Int. J. Man-Machine Studies, Vol.29, p579-587

Gruber, T., 1991, 'Learning why being told what', IEEE Expert, Vol.6, No. 4, p65-75

Gugerty, L., 1993, 'The use of analytical models in human-computer interface design', Int. J. Man-Machine Studies, Vol.38, p625-660

Guggenheim, L.J. and Whitfield, T.W., 1991, 'Designer integration in CIM: a motivational approach', Int. J. Man-machine Studies, Vol.35, p455-466

Guidon, R., 1990, 'Knowledge exploited by experts during software systems design', Int. J. Man-Machine Studies, Vol.33, p279-304

Guignat, G., 1993, 'Vital signs monitored with GUIs', Computing Canada, Vol. 19, No. 6, p31

Hand, D.J., 1985, 'Chapter 5: Computer assisted diagnosis', Artificial intelligence and psychiatry, New York: Cambridge Uni. Press, p137-167

Hart, P.E., 1982, 'Directions for AI in the Eighties', SIGART Newsletter, Vol79, p11-18

Hayes-Roth, F. and Jacobstein, N., 1994, 'The state of knowledge-based systems', Communications of the ACM, Vol. 37, No.3, p27-39

Hewett, T.T., 1986, 'The role of iterative evaluation in designing systems for usability', People and Computers: Designing for usability, (eds) Harrison, M.D. and Monk, A., Cambridge, UK: Cambridge Uni. Press

Heydemann, M., Hoffmann, R. and Schmidt, R., 1991, 'Interference among text-editing commands: Fan-effects and the role of system consistency', Acta. Psychologica, Vol.78, p259-285

Hilden, J. and Habbema, J.D.F., 1990, 'Evaluation of clinical decision aids - more to think about', Med. Inform., Vol.15, No.3, p275-284

Hoffman, R.R., 1987, 'The problem of extracting the knowledge of experts from the perspective of experimental psychology', AI Magazine, Vol.1, No.2, p35-48

Horrocks, J.C., McAdam, W.A.F., Devroede, G., Gunn, A.A. and Zoltie, N., 1976, 'Some practical problems in transferring computer-aided diagnosis from one geographical area to another', Decision-making and Medical Care, (eds) de Dombal, F.T. and Gremy, F., North-Holland: Elsevier Science

Howarth, I., 1987, 'Psychology and Information Technology', Information Technology and People, (ed) Blackler, F. and Oborne, D., Pub. Brit. Psychological Soc.

Hripcsak, G., 1991, 'Arden syntax for medical logic modules', M.D. Computing, Vol.8, No.2, p76-78

Hudson, D.L. and Cohen, M.E., 1985, 'Role of the user interface in a medical expert system', Comp. Appl. in Medical Care, (ed) Ackerman, M.J., Los Angeles, Calif.: IEEE Computer Soc. Press, p232-236

Innes, F.T., Goldfarb, N.I. and Balaban, D.J., 1985, 'Development of a family medicine clinical encounter database system', Computer applications in medical care, (ed) Ackerman, M.J., Los Angeles, Calif.: IEEE Comp. Soc. Press, p123-129

Innocent, P.R., 1982, 'Towards self-adapting interface systems', Int. J. Man-Machine Studies, Vol.25, p33-51

Irgon, A., Zolnowski, J., Murray, J. and Gersho, M., 1990, 'Expert systems development: A retrospective view of five systems', IEEE Expert, Vol.5, No.3, p25-39

Jackson, S.A., 1988, 'Evaluation of Knee Ligament Injuries', Final year B.Sc. (Hons) Computer Science project, Leicester Polytechnic

Jacob, V.S., Gaultney, L.D. and Salvendy, G., 1986, 'Strategies and biases in human decision-making and their implications for expert systems', Behaviour and Information Technology, Vol.5, No.2, p119-140

Jefferies, R., Turner, A.A., Polson, P.G. and Atwood, M.E., 1980, 'The process involved in designing software', Cognitive skills and their acquisition, (ed) Anderson, J., San Francisco: W.H. Feeman, p255

Johnson, P.E., 1983, 'What kind of expert should a system be ?', J. Medicine and psychology, Vol.8, P77-97

Johnson, P. and Johnson, H., 1988, 'Practical and theoretical aspects of human-computer interaction', J. Info. Tech., Vol.3, No.3, p147-161

Jojo, L.P., 1994, 'Experiences with an expert system prototyping methodology', Expert Systems, Vol. 11, No.1, p13--21

Kahnemann, D. and Tversky, A., 1972, 'Subjective probability: A judgement of representativeness', Cognitive Psychol. No.3, p430

Kantorowitz, E. and Sudarsky, O., 1989, 'The adaptable user interface', Communications of the ACM, Vol.32, No.11, p1352-1358

Kassirer, J.P. and Gorry, G.A., 1978, 'Clinical problem solving: a behavioural analysis', Annuals of Internal medicine, Vol.89, p245-255

Kelly, G.A., 1955, The psychology of personal constructs, Norton: New York

Keyes, J., 1989, 'Why expert systems fail', AI Expert, Nov.1989, p50-53

Keyes J., 1990, 'Where's the expert is expert systems ?', AI Expert, March 1990, p61- 64

Kidd, A.L., 1987, 'Knowledge Acquisition - An introductory Framework', Knowledge acquisition for expert systems: a practical handbook, Kidd A.L. (ed), Plenum Press, p1-16

Klein, G.A., Calderwodd R. and MacGregor D., 1989, 'Critical Decision Method for elicitating knowledge', IEEE Trans on Sys., Man, and Cyb. Vol.19, No.3, p462-472

Kleinmuntz B., 1984, 'Diagnostic problem solving by computer', Computers in biology and medicine,Vol. 14, No. 3, p255-70

Kolodner, J.L., 1983, 'Towards an Understanding of the role of experiences in the evolution from novice to expert', Int. J. Man-Machine Studies, Vol.19, No.5, p497-518

Konner, M., 1993, The Trouble with Medicine, London: BBC Books

Koran, 1975, 'The reliability of clinical methods, data and judgements', New England J. Med., Vol 293, p226-228

Kuipers, B. and Kassirer J.P., 1984, 'Causal reasoning in medicine: analysis of a protocol', Cognitive Science, Vol.8, p363-385

Kunz, J.C., 1984, 'Use of Artificial Intelligence and simple mathematics to analyse a physiological model', PhD dissertation, Ann Arbor, Mi., University Microfilms International Dissertation Information Services

LaFrance, M., 1987, 'The knowledge acquisition grid: a method for training knowledge engineers', Int. J. Man-Machine Studies, Vol.26, p245-255

Langlotz, C.P., Fagan, L.M., Tu, S.W., Sikic, B.I. and Shortliffe, E.H., 1990, 'A therapy planning architecture that combines decision theory and artificial intelligence techniques', Readings in Uncertain Reasoning, (ed) Shafer, G. and Pearl, J., San Mateo, Calif.: Morgan Kaufmann, p214-238

Lea, R-J. and Chung, C-G, 1990, 'Rapid prototyping from structured analysis: executable specification approach', Info. Software Tech., Vol.32, No.9, p589-597

Leaper, D.J., Horrocks, J.C., Staniland, J.R. and de Dombal, F.T., 1972, 'Computer assisted diagnosis of abdominal pain using estimates provided by clinicians', Br. Med. J., Vol.4, p350-354

Lenat, D.B. and Feigenbaum, E.A., 1991, 'On the thresholds of knowledge', Artificial Intelligence, Vol.47, p185-250

Lesmo, L., Saitta, L. and Torasso, P., 1989, 'Dealing with uncertain knowledge in medical decision-making: A case study in hepatology', Art. Int. in Medicine, Vol.1, p105-116

Lewis, C. and Mack, R., 1982, 'Learning to use a text processing system: evidence from 'thinking aloud' protocols', Proc. Conf. Human Factors In Comp. Sys., p387-392

Lindley, D.V., 1985, Making Decisions, 2nd Edition, London: John Wiley and Sons

Lings, B., Adablah, N., Foster, L., Jin, Y. and Narayanan, A., 1991, 'On a human-centred approach to database systems', AI and Soc., Vol.5, p128-141

Lipscombe, B., 1989, 'Expert systems and computer-controlled decision-making in medicine', AI and Soc., Vol.3, p5-18

Lundsgaarde, H.P., 1987, 'Evaluating medical expert systems', Soc. Sci. Med., Vol.24, No.10, p805-819

Luqi, 1989, 'Software evolution through rapid prototyping', Computer, May, Vol.22, No.5, p13-25

Lusted, L.B., 1976, 'Clinical decision-making', Decision-making and Medical Care, (eds) de Dombal, F.T. and Gremy, F., North-Holland: Elsevier Science

Lynn, M.P. and Bockanic, W.N., 1994, 'Legal liability of the knowledge engineer', J. Systems Management, May, p6-10

MacLean, A., Barnard, P.J. and Wilson, M.D., 1985, 'Evaluating the human interface of a data entry system: user choice and performance measures yield different tradeoff functions', People and Computers: Designing the interface, (ed) Johnson, New York: Cambridge Press, p172-186

Mager, C.E., 1983, 'An experimental evaluation on on-line help for non-programmers', Proc. CHI'83-Human Factors in Computing Systems, Association for Computing Machinery, New York, p277-281

Mayer, R.E., 1981, 'The psychology of how novices learn computer programming', ACM Computer Surveys, Vol.13, p121-141

McDermid, J., 1987, 'The role of formal methods in software development', J. Info. Tech., Vol.2, No.3, p124-133

McDonald, C.J., 1976, 'Protocol-based computer reminders; the quality of care and the non-perfectibility of man', New England J. of Medicine, Vol.295, p1351-1355

Meadow, C.T., 1970, Man-Machine Communications, London: John Wiley and Sons

Michon, J.A., 1972, 'Processing of temporal information and the cognitive theory of time experineces', Proc. of 1st Conf. of Int. Soc. for the Study of Time, (eds) Fraser, J.T., Heber, F.C., Mueller, G.H., Springer-Verlag, Berlin

Miller, G.A., 1956, 'The magic number +7 or -2', Psychological Review, Vol.63, p81- 97

Miller, G.A., 1990, 'Trends and debates in cognitive psychology', Issues in cognitive modelling, (ed) Aitkenhead, A.M., Slack, J.M., London: Lawrence Erlbaum Associates Ltd., p3-13

Miller, I. and Thomas, J., 1976, 'Behavoural issues in the use of interactive systems: part I General issues', Int. J. Man-Machine Studies, Vol.9, No.5, p509-536

Miller, P.L. and Sittig, D.F., 1990, 'The evaluation of clinical decision support systems: what is necessary versus what is interesting', Medical Informatics, Vol.15, p165-190

Miller, R.B., 1968, 'Response times in computer conversations', Proc. Fall Joint Computer Conference, New York: AFIPS Press

Millington, D., Tulloch, I.M. and Gray, E.M., 1991, 'Current use of software tools in the development of medical and health services computer applications - a survey', J. Information Technology, Vol.6, p86-93

Milner, P.M., 1966, Physiological psychology, New York: Holt, Rinehart and Winston

Minsky, M., 1985, The society of the mind, New York: Simon and Schuster

Moidu, K. and Wigertz, O., 1989, 'Computers and physicians - an appraisal study', Med. Inform., Vol.14, No.1, p63-70

Montazemi, A.R., 1991, 'The impact of experience on the design of the user interface', Int. J. Man-Machine Studies, Vol.34, p731-749

Moore, P.G. and Thomas, H, 1976, The anatomy of decisions, Middlesex, England: Penguin Books

Mouradain, W., 1990, 'Knowledge acquisition in a medical domain' in AI Expert, Vol.5, No.7, p34-38

Mumford, E., 1988, 'Participative design for expert systems', 4th Int. Exp. Sys. Conf., p1-7

Mumford, E., 1991, 'Participation in systems design: what can it offer ?', Human Factors for Information Usability, (eds) Shackel, B. and Richardson, S., New York: Cambridge Uni Press

Murray, G.D., 1990, 'Assessing the clinical impact of a predictive system in severe head injury', Med. Inform., Vol.15, No.3, p269-273

Mykytyn, K. and Mykytyn, P., Lunce, S., 1993, 'Expert Identification and Selection: Legal Liability Concerns and Directions', AI and Soc., Vol.7, p225-237

Nabney, I. and Grasl, O., 1991, 'Combining AI and statistical methods to explore a database', Expert System Applications, August, p9-14

Neale, I.M., 1988, 'First generation expert systems: a review of knowledge acquisition methodologies', The Knowledge Engineer Review, Vol. 3, No.2, p105-145

Newell, A. and Simon, H.A., 1972, Human problem solving, Englewood Cliffs, New Jersey: Prentice Hall

Nickerson, R.S., 1981, 'Why interactive computer systems are sometimes not used by people who might benefit from them.', Int. J. Man-Machine Studies, Vol.15, p469-483

Nisbett, R.E. and Wilson, T.D., 1977, 'Telling more than we know: verbal reports on mental processes', Psychological Review, Vol.84, p231-259

Norman, D.A., 1988, The psychology of everyday things, Basic Books: New York

Norman, D.A., 1990, 'Twelve issues for cognitive science', Issues in cognitive modelling, (ed) Aitkenhead, A.M., Slack, J.M., London: Lawrence Erlbaum Associate Ltd, p309-337

Noyes, J., 1983, 'The QWERTY Keyboard: a review', Int. J. Man-Machine Studies, Vol.18, p265-281

Nykanen, P., 1989, 'Guidelines for the evaluation of computer- based support to clinical decision-making', SYDOL-5 Working group, Scandinavian Conf. on AI, (eds) Jaakkola H. and Linnainmaa, S., IOS: The Netherlands, p291-298

Oborne, D.J., 1985, Computers at work: A behavioural approach, Chichester: John Wiley and Sons

Oborne, D.J., 1987, 'Ergonomics and new technologies', Information technology and people, (ed) Blackler, F., Oborne, D., Brit. Psychological Society Publ., p65-85

Ogden, W. and Kaplan, C., 1986, 'The use of 'and' and 'or' in a natural language computer interface', Proc. Human Factors Soc. 30th Annual Meeting Santa Monica, CA: Human Factors Soc., p829-833

Olson, J.R. and Rueter, H.H., 1987, 'Extracting expertise from experts: methods for knowledge acquisition', Expert Systems, Vol.4, No.3, p152-168

Olson, J.R. and Nilsen, E., 1988, 'Analysis of the cognition involved in spreadsheet software interaction', Human-Computer Interaction, Vol.3, p309-350

Olson, J.R. and Olson, G.M., 1990, 'The growth of cognitive modeling in human-computer interaction since GOMS', Human-computer interaction, Vol.5, p221-265

Parsaye, K., Chignell, M., Khoshafian, S. and Wong, H., 1989, Intelligent Databases, New York: John Wiley and Sons

Parsaye, K. and Chignell, M., 1993, Intelligent Database tools and applications, New York: John Wiley and Sons

Parsons, S., 1994, 'Uncertainty in artificial intelligence', The Knowledge Engineering Review, Vol.9, No.1, p65-66

Pauker, S.G., Gorry, A., Kassirer, J.R. and Schwartz, W.B., 1976, 'Towards the simulation of clinical cognition', Amer. J. of Medicine, Vol.60, p981-996

Piaget, J., 1950, The psychology of intelligence, London: Routledge and Kegan Publ.

Piatetsky-Sharpiro, G., 1994, 'Knowledge discovery in databases: Progress report', The Knowledge Engineering Review, Vol.9, No.1, p57-60

Pollard, P. and Evan, J.St.B.T., 1983, 'The role of representativeness in statistical inference: a critical appraisal', Thinking and Reasoning: Psychological approaches, (ed) Evan, J.St.B.T., London: Routledge and Kegan Publ.

Pringle, M., Robins, S. and Brown, G., 1984, 'Computers in the surgery: The patient's view', Brit. Med. J., Vol.288, p289-291

Ratcliff, B., 1987, Software engineering: principles and methods, London: Blackwell Scientific Publ.

Rector, A.L., 1989, 'Helping with a humanly impossible task: integrating knowledge based systems into clinical care', Scandinavian Conf. on Artifical Intelligence '89, (ed) Jaakkola, H., Linnainmaa, S.,p453-465

Reggia, J.A. and Tuhrim, S., 1985, 'An introduction to computer-assisted medical decision-making II', M.D. Computing, Vol.2, No.4, p40-46

Rettig, M., 1993, 'Cooperative software', Communications of the ACM, Vol. 36, No. 4, p23-28

Roberts, H.J., 1990, 'Expert systems clubs: Design methods', The Computer J., Vol.33, No.6, p556-561

Rosch, E. and Lloyd, B.B., (eds), 1978, Cognition and categorisation, New Jersey: Erlbaum Hillside

Rossi-Mori, A. and Ricci, F.L., 1988, 'On the assessment of medical expert systems', Expert Systems and Decision Support in Medicine, (eds), Rienhoff, O., Piccolo, U., Schneider, B., Berlin: Springer, p292-297

Rossi-Mori A., Pisanelli, D.M. and Ricci, F.L., 1990, 'Evaluation stages and design steps for knowledge-based systems in medicine', Med. Infor., Vol.15, No.3, p191-204

Rumelhart, D.E. and Norman, D.A., 1990, 'Representation of knowledge', Issues in Cognitive Modelling, (eds) Aitkenhead, A.M. and Slack, J.M., London: Lawrence Erlbaum Associates Ltd.

Safran, C., Rury, C., Rind, D.M. and Taylor, W.C., 1991, 'A computer-based outpatient medical record for a teaching hospital', M.D. Computing, Vol.8, No.5, p291-299

Sandberg, J. and Wielinga, B., 1991, 'How situated is cognition ?', 12th International Joint Conference on Artificial Intelligence, (ed) Mylopoulos, J. and Reiter, R., San Mateo, Calif.: Morgan Kaufmann

Santhanam, R., 1993, 'Neither novice nor expert: the discretionary user of software', Int. J. Man-machine Studies, Vol.38, p201-229

Savage, L.J., 1972, The foundations of statistics, Dover: New York

Schank, R.C., 1990, 'Reminding and memory organisation', Issues in cognitive modelling, (eds) Aitkenhead, A.M. and Slack, J.M., London: Lawrence Erlbaum Associates Ltd., p229- 251

Schwartz, S. and Griffin, T., 1986, Medical thinking: The psychology of medical judgement and decision-making, Springer: New York

Shackel, B., 1986, 'Ergonomics in design for usability', People and Computers: Designing for usability, (eds) Harrison, M. and Monk, A., Cambridge, UK: Cambridge Uni. Press

Shaw, M.L.G., 1980, 'Acquisition of personal knowledge', Int. J. Policy Analysis and Information Systems, Vol.4, No.4, p401-429

Shaw, M.L. and Woodward, J.B., 1988, 'Validation in a knowledge support system: construction and consistency with multiple experts', Int. J. Man-Machine Studies, Vol.29, p329-350

Shneiderman, B., 1987, Designing the user interface: strategies for effective human-computer interaction, New York: Addison-Wesley

Shneiderman, B., 1991, 'A taxonomy and rule base for the selection of interaction styles', Human Factors for Informatics Usability, (eds) Shackel, B. and Richardson, S.J., New York: Cambridge Uni. Press

Shiffman, S., Wu, A.W., Poon, A.D., Lane, C.D., Middleton, B., Miller, R.A., Masarie, F.E., Cooper, G.F., Shortliffe, E.H. and Fagan, L.M., 1991, 'Building a speech interface to a medical diagnostic system', IEEE Expert, Vol. 6, No. 1, p41-50

Shiffron, R.M. and Schneider, W., 1977, 'Controlled and automatic human information processing II perceptual learning, automatic attending and a general theory', Pyschological Review, Vol.84, p127-190

Shortliffe, E.H., Buchanan, B.G. and Feigenbaum, E.A., 1984, 'Knowledge engineering for medical decision-making: a review of computer-based clinical decision aids', Readings in medical Artificial Intelligence, (eds) Clancey, W.J. and Shortliffe, E.H., New York: Addison-Wesley, p35-71

Shumway, J.M., Jacknowitz, A.I. and Abate, M.A., 1990, 'Analysis of physicians', pharmacists', and nurses' attitudes towards the use of computers to access drug information', Meth. of Info. in Med., Vol.29, p99-103

Silverman. B.G., 1990, 'Critiquing human judgement using knowledge acquisition systems', AI magazine, Vol. 11, No. 3, p60-79

Simon, H.A., 1990, 'Information-Processing Theory of Human Problem Solving', Issuses in Cognitive Modelling, (ed) Aitkenhead, A.M. and Slack, J.M., London: Lawrence Erlbaum Associates Ltd., p253-278

Simpson, H.A., 1982, 'A human factor guide for programming design: taking the user into account in the design of software', Byte, Vol.7, p108-132

Sorgaard, P., 1991, 'Evaluating expert system prototypes', AI and Soc., Vol.5, p3-17

Southwick, R.W., 1991, 'Explaining reasoning: an overview of explanation in knowledge-based systems', The Knowledge Engineer Review, Vol.6, No1, p1-20

Sowa, J.F., 1984, Conceptual structures: Information processing mind and machine, Massachusetts: Addison Wesley

Spiegelhalter, D.J. and Knill-Jones, R.P., 1984, 'Statistical and knowledge-based approaches to clinical decision support systems, with an application to gastroentrology', J.R. Stat. Soc., Vol.147, p35-77

Staggers, N. and Norcio, A.F., 1993, 'Mental models: concepts for human-computer interaction research', Int. J. Man-machine Studies, Vol.38, p587-605

Steels, L., 1987, 'The deepening of expert systems', AICOM, Vol.0, No.1, p9-16

Sundstrom, G.A., 1993, 'User modelling for graphical design in complex dynamic environments: concepts and prototype implementations', Int. J. Man-machine Studies, Vol.38, p567-586

Sutcliffe, A.G., 1990, 'Integrating specifications of human computer interface with Jackson system development', Information and Software Technology, Vol.32, No.10, p665-676

Sutcliffe, A.G. and McDermott, M., 1991, 'Integrating methods of human-computer interface design with structured systems development', Int. J. Man-Machine Studies, Vol.34, p631-655

Sweeney, M., Maguire, M. and Shackel, B., 1993, 'Evaluating user-computer interaction: a framework', Int. J. Man-Machine Studies, Vol.38, p689-711

Szolovits, P. and Pauker, S.G., 1978, 'Categorical and probabilistic reasoning in medical diagnosis', Artificial Intelligence, Vol.11, p115-144

Tate, G., 1990, 'Prototyping: helping to build the right software', Info. Software Tech., Vol.32, No.4, p237-244

Taylor, T., Aitchison, J. and McGirr, E.M., 1977, 'Doctors as decision makers: a computer assisted study of diagnosis as a cognitive skill', Brit. Med. J., Vol.3, p35-40

Thomas, J.C., 1983, 'Psychological issues on the design of Database query languages', Designing for Human-Computer Communication, (eds) Sime, M.E. and Coombs, M.J., London: Academic Press, p173-209

Thomas, J.C., 1987, 'Database querying languages', Designing for human-computer communication, Series: Computers and People, London: Academic Press

Thomson, N., 1984, 'Human memory: Different stores with different characteristics', Fundamentals of Human-Computer Interaction, (ed) Monk, A., London: Academic, p49-63

Thomson, R., 1966, The psychology of thinking., Middlesex: Penguin Books

Todd, P. and Benbasat, I., 1992, 'The use of information in decision-making: an experimental investigation of the impact of computer-based decision aids', MIS Quarterly, Vol.16, No.3, p373-393

Trumbly, J.E., Arnett, K.P. and Martin, M.P., 1993, 'Performance effect of matching computer interface characteristics and user skill level.', Int. J. Man-Machine Studies, Vol.38, p713-724

Trumbly, J.E., Arnett, K.P. and Johnson, P.C., 1994, 'Productivity gains via an adaptive user interface: an empirical analysis', Int. J. Human-Computer Studies, Vol. 40, p63-81

Tuhrim, S. and Reggia, J.A., 1986, 'Feasibility of physician-developed expert systems', Med. Decis. Making, Vol. 6, p23-26

Tversky, A. and Kahnemann, D., 1971, 'Belief in the "law of small numbers" ',Psychol. Bull., No76, p105

Tversky, A. and Kahnemann, D., 1990, 'Judgment under uncertainty: heuristics and biases', Reading in Uncertain reasoning, (eds) Shafer, G. and Pearl J., San Mateo, Calif.: Morgan Kaufmann, p32-39

Valluy, L., Nykanen, P., Rantanen, J., Salmela, J., Gronroos, P. and Lumio, J., 1989, 'An approach to combine knowledge validation and evaluation with development in medical AI', Scandinavian Conf. on AI, (eds) Jaakkola H. and Linnainmaa, S., IOS: The Netherlands, p753-764

Van Hoe, R., Poupiyi, K., Vandierendonck, A. and De Soete, G., 1990, 'Some effects of menu characteristics and user personality on performance with menu-driven interfaces', Behaviour and Information Technology, Vol.9, No.1, p17-29

Vaubel, K.P. and Gettys, C.F., 1990, 'Inferring user expertise for adaptive interfaces', Human-computer interaction, Vol.5, p95-117

Vincent, R., 1993, 'Computers and the Clinician', The Computer Bulletin, Vol.5, No.3, p12-13

Wason, P.C. and Johnson-Laird, P.N., 1972, Psychology of reasoning: structure and content, Cambridge, Mass.: Harvard University Press

Wason, P.C., 1980, 'The verification task and beyond', The Social Foundations of Language and Thought, (ed) Olson, D.R., New York: Norton

Wasson, B. and Akselsen, S., 1992, 'An overview of on-line assistance: from on-line documentation to intelligent help and training', The Knowledge Engineering Review, Vol.7, No.4, p289-322

Waterman, D.A., 1986, A guide to expert systems, Massachusetts: Addison-Wesley

Watts, R.L., 1987, 'Measuring Software Quality', NCC Publications, Manchester

Weitzel, J.R. and Kerschberg, L., 1989, 'Developing knowledge-based systems reorganising the system development life-cycle', Communications of the ACM, vol.32, No.4, p482-488

Welbank, M., 1987, 'Knowledge acquisition: A survey and British Telecom experience', Proc. First European Workshop on Know. Acq. for Knowledge-based sys., (ed) Addis, T., Boose, J. and Gaines, B., Reading Uni. Reading

Wennberg, J. and Gittelsohn, A., 1969, 'Variation in medical care among small areas', Scientific American, Vol.246, p120-134

White, A.P., 1985, 'Inference deficiencies in rule-based expert systems', in Research and Development in Expert Systems, (ed) Bramer, M., Cambridge, UK: Cambridge Univ. Press.

Whiteside, J., Bennet, J. and Holtzblatt, K., 1987, 'Usability engineering: our experience and evolution', Handbook of human-computer interaction, (ed) Helander, M., North-Holland: Elsevier Science

Wielinga, B.J. and Breuker J.A., 1985, 'Interpretation of verbal data for knowledge acquisition', Advances in Artificial Intelligence, (ed) O'Shea, T., North-Holland: Elsevier Science, p3-14

Wielinga, B.J and Breuker, J.A., 1987, 'Models of Expertise', Advances in Artificial Intelligence II, (ed) Du Boulay, B., Hogg, D. and Steels, L., North-Holland: Elsevier Science

Williges, R.C., Williges, B.H. and Elkerton, J., 1987, 'Software interface design', Handbook of human factors, (ed) Salvendy, G., New York: John Wiley and Sons

Wilson, M., Duce, D. and Simpson, D., 1989, 'Life-cycles in software and knowledge engineering: a comparative review', The Knowledge Engineer Review, Vol.3, No.4, p189-204

Woods, D.D, 1986, 'Cognitive technologies: The design of joint human-machine cognitive systems', AI Magazine, Winter, Vol.6, p86-92

Wright, P. and Lickorish, A., 1994, 'Menus and memory load: navigation strategies in interactive search tasks', Int. J. Human-Computer Studies, Vol. 40, p965-1008

Wyatt, J., 1987, 'The evaluation of clinical decision support systems: a discussion of the methodology used in the ACRON project', AIME 87, (eds) Fox, J., Fieschi, M. and Engelbrecht, R., Berlin: Springer Verlag, p15-24

Wyatt, J. and Spiegelhalter, D., 1990, 'Evaluating medical expert systems: what to test and how ?', Med. Inform., Vol.15, No.13, p205-217

Yazdani, M., 1989a, 'Shells versus toolkits', Expert System User, Vol.5, No.1, p13

Yntenna, D.B. and Torgenson, W.S., 1961, 'Man-computer cooperation in decsions requiring common-sense', Decision-making, (ed) Edwards, W. and Tversky, A., Middlesex: England, Penguin Books, p300-315

Young, J., 1989, 'Human-centred knowledge based systems design', AI and Soc., Vol.3, No.2, p80-88

Yourdan, E., 1986, Managing the structured techniques: strategies for software development in the 1990's, Yourdan Press Inc.: New York

Zhang, X.J. and Yao, J.T.P., 1988, 'Automation of knowledge organisation and acquisition', Microcomputers in Civil Engineering, Vol.3, p1-12

Appendix A

Knowledge Acquisition Tools and Techniques

Computer Science and Artificial Intelligence (AI) are not the only fields that have needed to acquire knowledge from humans. There are the areas of Psychology, Philosophy, Sociology and Anthropology to list just a few. Data and information gathering from various sources have consequently been major topics. However, lessons that could have been learned from the experiences of these other disciplines were ignored by the computer fields, 'knowledge elicitation is treated by many involved with it as if it was a new problem, unrelated to work in other fields. This sort of belief leads to the reinventing of wheels. In particular, the work of the ergonomists and applied psychologists have been confronting the central problem of knowledge elicitation for decades' (Diaper, 1989). Therefore, although these investigations resulted in a large body of literature, which was also relevant to the knowledge acquisition problems confronting AI researchers, it has only been fairly recently that reviews of the other disciplines have been undertaken.

One example of a knowledge acquisition technique taken from another scientific discipline is the ethnographic technique that was developed by anthropologists. Forsythe and Buchanan (1989) proposed using this technique to overcome the difficulty that during the early stages of developing a system, knowledge engineers do not know which questions to ask nor in fact the answers they require.

Anthropology, in the area of problem definition and methodology, is concerned with researching and investigating how humans understand, organise, process and use symbolic information. The ethnographic methodology combines direct and non-direct interviewing with observations and the use of documentary material. It was developed to enhance information gathering during the informal and unstructured situations often encountered by anthropologists.

Consequently, AI researchers are now beginning to use skills and techniques gleaned from other disciplines to attempt to overcome a number of the difficulties associated with the knowledge acquisition stage. However, most of the techniques used still rely on old psychological methods, for example interviewing and protocol analysis, even many of the machine-aided knowledge acquisition tools are based on these approaches. The following sections briefly review a number of the commonly used methods for gathering information.

a) Interviewing

In the past, interviewing has been the main methodology used to develop expert systems. It is a good method for acquiring the general structure of a domain and the terminology used by experts but it has its limitations. It is poor at eliciting problem-solving strategies used by an expert; the context in which the rules are true; or, the weights of evidence for probabilistic reasoning (Welbank, 1987).

As described previously, possessing rules and heuristics and being able to use them does not mean that an expert can express the rules verbally. Yet it is this knowledge that must be explicit before it can be programmed into a computer. Berry and Broadbent (1984) carried out an experiment to demonstrate this problem. Subjects were taught complex rules concerning sugar production in a computer simulation game, by enabling them to control the program to achieve a required result. When subjects were given more extensive training however they did not show their superior knowledge in questionnaires conducted after the experiment. They could only demonstrate their advantage through playing the game.

Even if experts can verbalise their knowledge, there is not necessarily a correlation between the verbal reports and the actual way experts behave in practise (Wason and Evans, 1975; Bainbridge, 1986). Nisbett and Wilson (1977) discuss this problem when experts, during retrospection, construct ad- hoc theories to account for their behaviour. As Evans (1988) points out, 'not only may people not know their own mental processes but they may think they do and produce a misleading report.'

Other problems exist with the biases and errors created by the way the interviewer phrases questions (LaFrance, 1987) and the interpretation of responses. Detecting and avoiding errors of judgement is the responsibility of the interviewer. To be able to do this however, the interviewer must be able to recognise the errors and this requires substantial knowledge of the domain in question.

Forsythe and Buchanan (1989) outline other pitfalls that the knowledge engineer must negotiate in order to carry out a successful interview - successful in terms of the knowledge acquired and the important bond established between the expert and the knowledge engineer. A number of these difficulties arise from the ease of use and the flexibility of interviewing. For example, as interviewing uses the medium of conversation, which is an everyday activity, it has been reported that knowledge engineers sometimes assume that knowledge acquisition is merely a matter of 'chatting' with an expert.

Interviewing is not that easy. It is very easy, however, to interview badly with the likely result of learning little and the high possibility of alienating an expert. Much relies on the interviewer - 'interviewing does not just happen: the knowledge engineer must make it happen. Far from coming naturally, interviewing is a difficult task that requires planning, stage-management technique and a lot of self-control' (Forsythe and Buchanan, 1989).

Such is the importance of getting the balance correct, between obtaining the required information and acknowledging the expert not only as a person but as a professional, that a number of practical guides for interviewing have been written to aid knowledge engineers, for example, - Cordingley (1989), Forsythe and Buchanan (1989), Davies and Hakiel (1988) and Olson and Reuter (1987). These highlight many of the pitfalls and dangers of interviewing and suggest possible solutions or methods of avoiding awkward situations in the first place.

In spite of these problems, interviewing experts can provide a considerable amount of information about the basic concepts of the domain, the nature of the problem and the goals of decision processes (Evans, 1988). Gammack and Young (1984) agree, 'it quickly generates a lot of knowledge that indicates the terminology and main components of the domain. Thus it has an important role to play early on in the process of knowledge elicitation in order to get some basic concepts and information established as a framework for what comes later.'

As Welbank (1987) points out, it can also be quite successful at eliciting poorly remembered details by asking the same question many times in different ways, as well as revealing covert knowledge by asking for examples rather than rules.

Neale (1988) describes and discusses sixteen different interviewing strategies which stretch from:

Structured interviews - involving detailed depth-first sequencing of topics to elicit all knowledge relating to a particular concept or model,

through

Tutorial interviews - the expert prepares an introductory talk,

and

Teachback interviews - the expert describes a procedure to the interviewer, who then 'teaches' it back in the expert's terms and to the expert's satisfaction,

to

Forward scenario simulation - the expert describes in detail how a hypothetical case would be solved.

Greenwell (1988), Shadbolt (1988), Gordon (1989) and Cordingley (1989) are other authors who have recently reviewed the available interviewing techniques in some depth. The number of possible interviewing methods and the amount of in-depth literature on the subject further confirms that interviewing a human expert is not as straight-forward or as easy as it may first appear.

Consequently, careful planning for the interviews to avoid the reported pitfalls and thorough analysis of the collected information must be carried out. Since the quality of the data gathered and the time taken to collect it have direct implications on the system being developed, closer scrutiny must be placed on the interviewing technique adopted and the methods followed by the knowledge engineer - 'if the information that goes into a knowledge base is poorly understood or incomplete, the most sophisticated representation or inference schemes will not produce a good system' (Forsythe and Buchanan, 1989).

b) Protocol Analysis

This procedure is based upon individuals being given real or simulated tasks to perform and asking them to verbalise their thoughts and actions whilst they work. This can be recorded on tape and later a 'protocol' can be produced from the recorded details. This protocol is then analysed for meaningful associations. Byrne

(1983) showed that these protocols need not be verbal but this is still the most common approach. The major benefit of protocol analysis is that in a more realistic situation an expert may reveal the knowledge used in problem-solving - particularly the heuristics which they are unable to articulate in an interview (Gammack and Young, 1984).

Ericsson and Simon (1980,1984) emphasise the importance of collecting verbal reports concurrently rather than retrospectively. They showed that information should be interpreted as displaying the results of cognitive processes rather than self-reported descriptions obtained from the introspective approach. It has been argued therefore that verbal reports collected in this manner provide details about the information heeded to by the expert at a given point and reflect the current contents of the short-term memory (Evans, 1988; Ericsson and Simon, 1984).

Due to the limitations of short-term memory, however, the verbal reports are often incomplete. This is because under heavy cognitive load subjects have to stop verbalising otherwise some of the limited short-term memory would be occupied and would result in an altered performance.

In addition, it has been suggested that asking experts to 'think aloud' is likely to make them approach the task in a different, more systematic way. This can result in changes to the underlying thought processes (Berry, 1987; Kassirer et al, 1982).

However, carrying out the analysis retrospectively leads to other problems. 'Subjects often make inaccurate or misleading inferences about their own thought processes. Observations about previous thought processes are particularly subject to retrospective biases, because judgements are significantly influenced by knowledge of the outcomes.' (Kassirer et al, 1982).

Byrne (1983) agrees that allowing subjects to analyse their own protocols in retrospect invites rationalisations associated with introspection. This can lead to inaccuracies, errors and biases in the knowledge acquired.

Nisbett and Wilson (1977) carried out the first major study to review whether there was any relation between verbal reports and the actual cognitive processes used by an expert. They concluded that:

- people are often unable to identify the existence of evaluative or motivational responses,
- people have difficulty in reporting that a process has occured, e.g. interviewing a creative artist about his/her creative cognitive processes,
- people often have difficulty recognising the existence of critical stimuli,
- even if the stimulus and response is known, people cannot accurately explain the relationship between them.

Bainbridge (1979) identified another problem with protocol analysis. Experts may not verbalise what is 'obvious' to them, i.e. 'common-sense' knowledge and the knowledge that experts, with many years experience, have 'compiled' into a single association. This problem of incompleteness is heightened further by the fact that certain types of thinking lead to little or no verbal expression (Welbank, 1983; Bainbridge, 1986; Berry, 1987).

Therefore though the structure of an expert's problem-solving method may be discovered, conclusions about the limit of that expert's knowledge cannot be inferred (Kuipers and Kassirer, 1983) nor can it be assumed that a reasoning process described is complete (Kassirer et al, 1982).

In addition, there is the question of which methods should be used to analyse the protocols. These methods are interpretative and influenced both by an investigator's underlying beliefs regarding human behaviour and problem-solving and a knowledge of the actual problem domain.

To be effective a knowledge engineer must be sufficiently acquainted with a domain to understand an expert's task - 'we have encountered instances in which an ambiguous statement could be clarified, or a tentative analysis rejected, only by a reader who was deeply familiar with the typical thought and language patterns of physicians' (Kassirer et al, 1982).

This knowledge of the problem domain is essential to ensure that the analysis undertaken is valid. If a simulated task is to be used then the selection of problems is crucial in order to get a truely representative sample. If a natural task is being observed then the behaviour must be recorded for a sufficiently long period of time to cover a representative sample of activities (Berry, 1987).

Protocol analysis is therefore labourious and very time-consuming (Garg-Janardan and Salvendy, 1987). Burton et al (1988) discovered that 'not only does protocol analysis take longer to perform and analyse than the comparable technique (interviewing) but it also seems to retrieve a substantially smaller amount of the necessary information than the other techniques (interviews, laddered grid, card sorting).'

An attempt to reduce the amount of time required was put forward by Breuker and Wieling (1984). They suggested that an expert should first select a number of representative problems. This should not only reduce the work involved but could also provide information on the way an expert classifies a domain.

One final warning concerns the selection of a domain for protocol analysis. The choice is crucial. It may be that protocol analysis is a particulary useful method for eliciting procedures that experts use in problem-solving but the action of 'thinking aloud' must not intrude into these problem-solving processes. As Olson and Rueter (1987) explain, 'if verbal information is produced while someone makes inferences to him/herself, or in identifying salient features of the objects in the situation then the information from the protocols is acceptable data. However, there are - tasks for which there is no natural verbalisation; perceptual-motor tasks are examples of these. Verbalisation of perceptual-motor tasks makes someone attend to aspects not normally attended to, and the attention required to report on the process usurps resources normally devoted to the task itself.' In these cases, they point out, the details gathered are likely to be distorted or even wrong.

c) Multidimensional Techniques

These techniques attempt to map an expert's representation by uncovering the criteria that are used by an expert to organise the concepts of a problem domain. Numerous techniques exist, for example, ordered trees from recall, laddered grids, matrix analysis. The three most common methods are briefly described below. Further details of these and others can be found in the reviews undertaken by Neale (1988), Shadbolt (1988), Olson and Reuter (1987) and Cordingley (1989).

Card Sorting

A set of cards each bearing a name of one concept are set out randomly. The expert is asked to sort the cards into groups according to a certain citerion. This separation illustrates one 'dimension' of how the expert classifies the concepts. The task is repeated for all the different ways the expert believes the concepts vary.

It is a good method for structuring a large set of concepts, so long as a natural hierarchical strucure exists in the domain. If there is no such structure than the results could be confused and of limited use (Gammack and Young, 1984)

Gammack (1987) believes that card sorting does have its merits, 'sorting is a task people find natural and easy, and not just concepts, but pictures, sentences or domain problems may be used as stimuli.'

Burton et al (1988), however, noted that the experts themselves were unenthusiastic about the task.

Multidimensional Scaling

There are a number of differing procedures, though generally each of the concepts is compared with all of the others and an estimate of their similarity is given.

This is a possible technique to use when the concepts vary over a small number of dimensions. As Gammack and Young (1984) noted, 'for expert knowledge elicitation, this technique seems appropriate when there are a number of closely related concepts, typically not well differentiated by novices, and expertise consists in being able to make discriminations.'

However, these techniques are considered as being complicated and rather strenuous on the expert (Burton and Shadbolt, 1988; Welbank, 1987). Gammack (1987) and Olson and Reuter (1987) agree that it is very demanding on the expert but it does provide an efficient classification of the problem space. This could be helpful if the expert is not aware of the structure of the problem domain or as a check that the structure given is in fact the one that the expert uses.

Cooke and McDonald (1987) make the point that scaling techniques do reduce the dependence on introspection and verbal reports, compared with the traditional interview and protocol analysis techniques.

Repertory Grid

The repertory grid technique is possibly best known in its automated form developed by Boose (1985,1986,1988). Kelly (1955) originally devised this technique to elict the constructs which people use to view the world. A list of objects within the domain is generated by the expert and the investigator asks the expert to name constructs on which the set of objects shows similarities or differences. The expert then ranks each object with respect to the construct on a scale generally 1 to 5. Statistical analysis of the grids by multivariate methods are carried out and these can reveal clustering or factoring of the objects and constructs.

Greenwell (1988) suggests that the number of objects selected for the analysis should be between 10 and 20. However, Welbank (1983) noted that the technique, when carried out manually, becomes unworkable with more than 10 objects. Consequently, this restriction limits the usefulness of this technique as most domains have many more objects than ten.

As a method for eliciting declarative knowledge, it is relatively straightforward and it resembles a highly structured interviewing technique. It asks an expert to access and verbalise not only the declarative knowledge but to also place a quantitative value on knowledge links. Therefore, it provides a quantitative index of the relationship between solution and trait (Gordon, 1989).

Hart (1986) states that 'the main assest of a grid is that it makes an expert think carefully about a problem.' Furthermore, she believes that after drawing up a grid an expert will probably be more aware of rules, i.e. an expert's perception may be clearer. This will greatly assist in further knowledge acquisition sessions.

Gordon (1989) however warns that caution must be exercised in the interpretation of the information elicited: 'There is no evidence that forcing experts to quantify all possible solution or trait pairs will yield information resembling how the information is stored in the human memory'.

Boose (1986) also points out that the repertory grid method has limitations - 'it is difficult to apply grid methodology to elicit causal knowledge, procedural knowledge or strategic knowledge.'

Evans (1988) agrees that it does not provide much information about the procedural knowledge which an expert possesses, though Gordon (1989) claims that it creates 'artificial' procedural knowledge - 'Unfortunately, at this time we do not know the extent to which this artificial procedural knowledge is similar to the expert's procedural knowledge. Research is needed to determine the extent to which this artificial creation of procedural knowledge decreases the effectiveness or expertise of the system.'

d) Machine-aided Knowledge Acquisition Tools

As Barfield (1986) states 'all these methods [above] involve people other than the experts themselves in the interpretation of the knowledge and expertise. Thus all these methods offer the potential for entering misinformation into the knowledge base.'

Shaw and Gaines (1983) also believe that less knowledge would be lost from an expert interacting directly with a computer than with using an intermediary whose lack of domain knowledge could be more destructive. An attempt, therefore, to find a more effective way for knowledge acquisition and also to save time and money has meant there has been a move towards machine-aided knowledge acquisition.

Kitto and Boose (1989) outline the necessity of choosing the correct tools for the type of application task. This is however not easy since it is possible for complex applications to require several problem-solving techniques to resolve the total problem and yet most current knowledge acquisition tools support only one problem-solving method. Rajan (1989) agrees that one technique does not currently exist that can elicit or encompass all the types of knowledge an expert may use.

Waterman (1986) classified expert systems tools into three categories : programming languages (such as PROLOG and LISP); knowledge engineering languages (such as shells and toolkits); and, system building aids (such as machine induction and knowledge acquisition tools). The next sections briefly review the last two categories. Programming languages are not considered as they are not really appropriate to this study as a user would be a naive computer user and therefore would not have the necessary computing knowledge.

Shells and Toolkits

There are merits to using shells, they permit faster and cheaper development but they constrain a designer into the limited formalisms that they support. This restriction becomes unacceptable if the knowledge to be represented is of any real complexity (Hayward, 1985).

Over the past few years, the drive has been to develop shells for the PC with the idea that business and other organisations could make use of the shells to develop their own expert systems. However, Gold (1986) noted that, 'most [of the shells] are beyond the technical proficiency of the average business user' and that 'most vendors recommended that the users have at least some experience of programming'. Forsythe (1987) agreed that the learning curve for a user was rather steep.

The advancement in shells to include additional facilities such as screen editors and mechanisms for representing uncertainty, has resulted in the distinction between shells and toolkits becoming blurred to such an extent that the terms are now often interchanged during use.

Toolkits vary a great deal in the features that they offer, their flexibility and their price. Gevarter (1987) carried out a detailed review of the available toolkits. Chung and Kingston (1989) compared the features of three commercial systems ART, KEE and Knowledge Craft.

One of the major difficulties that has emerged with toolkits is their complexity of operation. They tend to offer the knowledge engineer a bewildering cluster of features with little or no guidance as to the applicability of facilities for particular conditions (Breuker and Wielinga, 1987a; Mettrey, 1987).

In addition, as Neale (1988) concluded, although shells and toolkits offer a ready-made framework for building an expert system, they do impose various constraints. These restrictions could result in the expertise being inaccurately represented because the selected shell or toolkit uses the wrong knowledge engineering strategy, i.e. an inappropriate representation framework for the domain knowledge. Cooke and McDonald (1986) observed that to choose the knowledge representation scheme before acquiring the knowledge is like 'putting the cart before the horse'.

Machine Induction

Here an expert supplies a set of domain examples showing different types of decision. This is called a training set. Attributes which describe the examples are listed and the expert assigns values to those attributes. From this information the computer programme induces a set of rules, which are often constructed in the form of a decision tree (Neale, 1988).

However, no-one can be certain that an inductive inference is true unless all of the objects in the domain are known since induction is really a form of conjecture. Therefore, machine induction is only really suitable for fairly straight-forward, simple, well-defined problem domains as the training set has to cover all the possible cases, including unusual or 'difficult' cases (Mingers, 1987; Michalski and Chilansky, 1980). If it is incomplete or inadequate, the result is highly likely to be poor rules (Hart, 1986).

Neale (1988) asks, since experts can not account for all they know, how can a knowledge engineer be sure that a training set is adequate? How can one be sure of the attributes supplied and whether they constitute a sufficient set for the construction of a valid decision tree? Can an expert unambiguously assign values to these attributes?

Knowledge Acquisition Tools

Birmingham and Klinker (1993) identified the following two characteristics of specialised knowledge acquisition tools:

1) they presuppose a problem-solving method, as well as the structure of a knowledge base, i.e. they exploit a model of the expert system they generate;
2) they acquire knowledge, generate the acquired knowledge, check it for errors and then generate code.

McDermott (1988) and Gruber (1987) agree and continue stating that by presupposing the problem-solving method of the expert system that is generated or extended, the tools do not design a new knowledge base nor do they design a problem-solving method. Thus, they do not assist a user with performing detailed task analysis.

Therefore, as Birmingham and Klinker (1993) also point out, the narrow scope of knowledge acquisition tools result in two major drawbacks:

1) it is difficult to determine whether a tool is appropriate for a given application;
2) the tools break once they encounter situations that cannot be solved by their presupposed

problem-solving methods and accompanying knowledge representations, i.e. the tools are brittle. Boose (1989) and Westphal (1989) have given a comprehensive review of existing knowledge acquisition tools.

Workbench Systems

Recently a number of workbench systems that contain sets of tools to support a knowledge engineer have also been developed. For example, KADS (Breuker and Wielinga, 1987ab), KEATS (Motta et al, 1988) and KRITON (Diederich et al, 1987; Linster, 1989). Unfortunately, these system are complex to operate and require specialist knowledge for them to be utilised with any degree of efficiency and/or effectiveness. 'The more powerful systems become, the more difficult they are to use. Before users will be able to take advantage of the power of high functionality computer systems, the cognitive costs of mastering them must be reduced' (Fischer, 1992).

Fischer (1992) lists four problems of high functionality systems:

a) users do not know about the existence of tools,
b) users do not know how to access tools,
c) users do not know when to use tools,
d) users cannot combine, adapt, and modify tools according to their specific needs.

They are thus not appropriate for use by non-computer literate or novice computer users.

Furthermore, though these systems contain a number of tools, Kitto and Boose (1989) doubt that 'a single advisory system could direct and monitor knowledge base development under several knowledge acquisition tools, particulary where the tools rely on built-in domain knowledge'. They continue by stating that knowledge acquisition tools do not currently exist for certain applications or problem-solving methods. In addition, most existing knowledge acquisition tools can support only one problem-solving method when generally several are required to solve the problem task. Morik (1987) agrees, 'the model to be represented in the expert system is pre-supposed. The system supports the encoding of a given model but not the building of the model itself.' Hence the major limitation, which is '(shared by all automated approaches, including machine learning) is the limited expressiveness of the representation. The machine can only elicit knowledge in terms that have already been operationally defined' (Gruber, 1991).

e) Summary

All of the techniques described above have drawbacks. In some cases, it is the complexity of the approach or the tools or the analysis of the information gathered; in others, it is the lack of accuracy and completeness of the reports; but in all cases they suffer from two major problems:

a) the ability of a technique to acquire only one type of knowledge or problem-solving strategy,
b) the restriction of only being able to represent the knowledge within one pre-defined structure.

Thus, although attempts have been made to obtain the knowledge by combining the results from two techniques or by representing the knowledge in two structures within one system, neither have worked since no-one really knows how the different methods should be integrated. This, in part, returns to the problem of not understanding the human information processing strategy or the representations used or, in fact, how all the structures and knowledge fit and work together. As Diaper (1989) states, 'perhaps one of the major reasons that the collaborative computer science and psychology venture of AI has, to date, generally failed, is that the psychologists are a very long way from elucidating even the most basic properties of information processing associated with human cognition.' 'There is no doubt that this is principally due to the already mentioned, apparently intractable problems associated with our ignorance of the real psychology of human knowledge representations'.

Reference List

Bainbridge, L., 1979, 'Verbal reports as evidence of the process operator's knowledge', in Int. J. Man-Machine Studies, Vol.11, p411-436

Bainbridge, L., 1986, 'Asking questions and accessing knowledge', in Future Computing Systems, Vol.1, No.2, p143-151

Barfield, W., 1986, 'Expert-novice difference for software: implications for problem- solving and knowledge acquisition', in Behaviour and Technology, Vol.5, No.1, p15-31

Berry, D.C., 1987, 'The problem of implicit knowledge', in Expert Systems, Vol.4, No.3, p144-151

Berry, D.C., and Broadbent, D.E., 1984, 'On the relationship between task performance and associated verbalised knowledge', in Quarterly Journal of Experimental Psychology, Vol.36a, p209-231

Birmingham, W., and Klinker, G., 1993, 'Knowledge-acquisition tools with explicit problem-solving methods', The Knowledge Engineering Review, Vol.8, No.1, p5-25

Boose, J.H., 1985, 'A knowledge-acquisition program for expert systems based in Personal Construct psychology', in Int. J. Man-Machine Studies, Vol.23, p495-525

Boose, J.H., 1986, Expertise Transfer for expert system design, Elsevier: New York

Boose, J.H., 1988, 'Uses of repertory grig-centered knowledge acquisition tools for knowledge based systems' in Int. J. Man-Machine Studies, Vol.29, p287-310

Boose, J.H., 1989, 'A survey of knowledge acquisition techniques and tools', Knowledge Acquisition, Vol.1, No.1, p3-38

Brachman, R.J., Amarel, S., Engelman, C., Engelmore, R.S., Feigenbaum, E.A. and Wilkins, D.E., 1983, 'What are expert systems ?', in Building Expert Systems, (eds) Hayes-Roth, F., Waterman, D.A., and Lenat, D.B., Reading: Addison-Wesley, p31-57

Breuker, J.A., and Wielinga, B.J., 1984, 'Techniques for knowledge elicitation and analysis', in Report 1.5 Esprit Project 12, University of Amsterdam, Amsterdam, July

Breuker, J.A., and Wielinga, B.J., 1987a, 'Use of models in the interpretation of verbal data', Knowledge Acquisition for Expert Systems - A Pratical Handbook, Kidd, A. (ed), New York: Plenum Press

Breuker, J.A., and Wielinga, B.J., 1987b, 'Knowledge-acquisition as modelling expertise: the KADS methodology', Proc. 1st European Workshop on Knowledge Acquisition for Knowledge-based Systems, (ed) Addis, T., Boose, J. and Gaines, B., Reading: Reading Uni.

Burton, A.M., Shadbolt, N.R., Hedgecock, A.P. and Rugg, G., 1988, 'A formal evaluation of knowledge elicitation techniques for expert systems: Domain I', in Research and Development in expert systems IV, (ed) Moralee, D.S.,p136-145

Byrne, R., 1983, 'Protocol analysis in problem solving', in Thinking and Reasoning: Psychological approaches, (ed) Evans, J.St.B.T., London: Routledge and Kogan Paul, p227- 250

Chung, and Kingston, 1989, 'State of the Art: Knowledge-based system toolkits', Presented at 5th Annual Meeting on Expert Systems in Medicine, p1-11

Cooke, N.M., McDonald, J.E., 1986, 'A formal methodology for acquiring and representing expert knowledge', in Proc. IEEE, Vol.74, No.10, p1422-1430

Cooke, N.M., McDonald, J.E., 1987, 'The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems', Int. J. of Man-Machine St., Vol.26, p533-550

Cordingley, E.S., 1989, 'Knowledge elicitation techniques for knowledge based systems' in Knowledge elicitation, (ed) Diaper, D., Ellis Horwood Ltd., p 89-177

Davis, R., 1979, 'Interactive transfer of expertise acquisition of new inference rules', in Artificial Intelligence, Vol.12, p121-158

Diaper, D., 1989, 'Designing expert systems: from Dam to Beersheba', Knowledge elicitation principles, techniques and applications., (ed) Diaper, D., Ellis Horwood Ltd.

Diederich, J., Ruhmann, I., May, M., 1987, 'KRITON: A Knowledge-acquisition tool for expert systems', Int. J. Man-Machine Studies, Vol.26, No.1, p29-40

Ericsson, K.A., and Simon, H.A., 1980, 'Verbal reports as data', in Psychological Review, Vol.87, No.3, p216-254

Ericsson, K.A., and Simon, H.A., 1984, Protocol Analysis: Verbal reports as data, MIT Press: Cambridge, Massachusetts

Eshelman, L., Ehert, J., McDermott, J., Tan, M., 1987, 'MOLE: A tenacious knowledge acquisition tool', in Int. J. Man-Machine Studies, Vol. 26, p41-54

Evans, J.St.B.T., 1988, 'The knowledge eliciation problem: a psychological perspective', in Behaviour and Information Technology, Vol.7, No.2, p111-130

Fischer, G., 1992, 'Domain ortiented design environment', Proc. of 7th Knowledge-based Software Engineering Conf., IEEE Comp. Soc. Press, p204-213

Forsythe, D.E., 1987, 'Engineering Knowledge: an anthropological study of an artificial intelligence laboratory', presented to a meeting of the Soc. for Social Studies of Science, Worcester, MA, Nov. 1987

Forsythe, D.E., and Buchanan, B.G., 1989, 'Knowledge acquisition for expert systems: some pitfalls and suggestions', in IEEE Trans. Sys., Man, and Cyb., Vol.19, No.3, p435- 441

Gammack, J.G., and Young, R.M., 1984, 'Psychological techniques for eliciting expert knowledge', in Research and Development of expert systems, (ed) Bramer, M.A., Cambridge Uni. Press, Cambridge: UK.

Gammack, J.G., 1987, 'Different techniques and different aspects of declarative knowledge', in (ed) Kidd, A.L., Knowledge acquisition for expert systems: A practical handbook, Plenum Publ.

Garg-Janaradan, and C., Salvendy, G., 1987, 'A conceptual framework for knowledge elicitation', in Int. J. Man-Machine Studies, Vol.26, p521-531

Gevarter, W.B., 1987, 'The nature and evaluation of commercial expert system building tools', Computer, Vol.20, No.5, p24-41

Gold, J., 1986, 'Do-it-yourself expert system', Computer Decisions, Vol.18, No.2, p76-81

Gordon, S.E., 1989, 'Theory and methods for knowledge acquisition', in AI Applications, Vol.3, No.3, p19-30

Greenwell, M., 1988, Knowledge engineering for expert systems, Ellis Horwood Ltd.

Gruber, T.R., 1987, 'Acquiring strategic knowledge from experts' in Int. J. Man-Machine Studies, Vol.29, p579-587

Gruber, T., 1991, "Learning why being told what",IEEE Expert, Vol.6, No. 4, p65-75

Hart, A., 1986, Knowledge acquisition for expert systems, Routledge and Kogan Paul: London

Hayward, S.A., 1985, 'Is a decision tree an expert system ?', in Research and Development in Expert Systems, (ed) Bramer, M.A., Cambridge:Cambridge Uni. Press, p185-192

Kahn, G., Nowlan, S., McDermott, J., 1985, 'Strategies for knowledge acquisition', in IEEE Trans. Pattern Anal. and Machine Intell., P.AMI-7, Vol.5, p511-522

Kassirer, J.P., Kuipers, B.J. and Gorry, G.A., 1982, 'Towards a theory of clinical expertise', in The American Journal of Medicine, Vol.73, p251-259

Kelly, G.A., 1955, The psychology of personal constructs, Norton: New York

Kitto, C.M., and Boose, J.H., 1989, 'Selecting knowledge acquisition tools and strategies based on application characteristics', in Int. J. Man-Machine Studies, Vol.31, p149-160

Kuipers, B., and Kassirer J.P., 1983, 'How to discover a knowledge representation for causal reasoning by studying an expert physicians', in Proc. 8th Int. J. Conf. Art. Intell. Vol. 1, (ed) Bundy, A., William Kaufmann: Los Altos, Calif.

LaFrance, M., 1987, 'The knowledge acquisition grid: a method for training knowledge engineers' in Int. J. of Man-Machine Studies, Vol.26, p245-255

Linster, M., 1989, 'Towards a second generation knowledge-acquisition tool', Knowledge Acquisition, Vol.1, No.2, p163-184

Marcus, S., 1987, 'Taking backtracking with a grain of SALT', Int. J. Man-Machine Studies, Vol.26, No.4, p393-398

McDermott, J., 1988, 'Preliminary steps towards a taxonomy of problem-solving methods', Automating Knowledge acquisition for expetr systems, Marcus, S., (ed), Kluwer

Mettrey, 1987, 'An assessment of tools for building large knowledge-based systems', AI Magazine, Winter, p81-89

Michalski, R.S., and Chilansky, R.L., 1980, 'Knowledge acquisition by encoding expert rules versus computer induction from examples: a case study involving soybean pathology', Int. J. Man-Machine Studies, Vol.12, p63-87

Mingers, J., 1987, 'Expert systems - rule induction with statistical data', Journal of the Operational Research Soc., Vol.38, No.1, p39-47

Morik, K., 1987, 'Knowledge-acquisition and machine learning - the issue of modelling', Proc. of 1st European Workshop on Knowledge Acquisition for Knowledge-based Systems, (ed) Addis, T., Boose, J. and Gaines, B., Reading: Reading Uni.

Motta, E., Eisenstadt, M., Pitman, K., West, M., 1988, 'Support for knowledge acquisition in the knowledge engineer's assistant (KEATS)', Expert Systems, Vol.5, No.1, p6-28

Musen, M.A., 1988, 'Conceptual model of interactive knowledge acquisition tools', in EKAW '88, (ed) Boose, J.H., Gaines, B.R., Linster, M.

Neale, I.M., 1988, 'First generation expert systems: a review of knowledge acquisition methodologies' in The knowledge engineer review, Vol. 3, No.2, p105-145

Nisbett, R.E., and Wilson, T.D., 1977, 'Telling more than we know: verbal reports on mental processes', in Psychological Review, Vol.84, p231-259

Olson, J.R., and Rueter, H.H., 1987, 'Extracting expertise from experts: methods for knowledge acquisition', in Expert Systems, Vol.4, No.3, p152-168

Rajan, T., 1989, 'Knowledge acquisition for model based expert systems', in Expert System User, p18-21, May

Shadbolt, N., 1988, 'Knowledge elicitation - the key to successful ES', in Expert Systems User, Oct., p22-25

Shaw, M.L.G., and Gaines, B.R., 1983, 'A computer aid to knowledge engineering', in Brit. Comp. Soc. Workshop Series Proc. 3rd Technical Conf. BCSSG on Expert Systems, p263-271

Wason, P.C., and Evans, J.St.B.T., 1975, 'Dual processes in reasoning ?' in Cognition, Vol.3, p141-154

Waterman, D.A., 1986, A guide to expert systems, Massachusetts, USA: Addison-Wesley

Welbank, M, 1983, A review of knowledge acquisition techniques for expert systems, Ipswich: Martlesham Consultancy Services

Welbank, M., 1987, 'Knowledge acquisition: A survey and British Telecom experience', in Proc. First European Workshop on Know. Acq. for Knowledge-based sys., (ed) Addis, T., Boose, J. and Gaines, B., Reading Uni. Reading: Uk.

Westphal, C. and McGraw, K., (eds), 1989, Special Issue on Knowledge Acquisition, ACM SIGART, Vol.108

Appendix B

Interface Considerations

The widespread introduction of computers has resulted in an increase of people, from all walks of life, needing to interact with computers in their daily lives. A user interface is the medium through which this interaction takes place. Consequently, a design of an interface must reflect the type of user, the task and, ultimately, the environment in which an end-system is to operate.

a) Keyboards

Dvorak:

The Dvorak keyboard had all the vowels and the most used consonants on the second row. This was to enable 70% of common words to be typed from this row alone. In general, the arrangement of the keys was such that the left hand typed the vowels and the frequent consonants were typed by the right hand. The reasons, it was argued, were to enable a more even distribution of finger movements and a bias towards the right hand. Furthermore, it decreased the need for movement between the rows by 90% and allowed 35% of all words normally used to be typed from the middle row (Oborne, 1987).

Though some studies have claimed that little difference exists between the QWERTY and the Dvorak keyboards (Alden et al, 1976; Martin, 1972; and Dunn, 1971), all argue that the Dvorak is easier to learn, reduces likelihood of error and fatigue and increases the speed of entry. Yet the acceptance of the design has still been extremely slow. As Shneiderman (1987) points out this is 'an interesting example of how even documented improvements are hard to disseminate because the perceived benefit of change does not appear to outweigh the effort [involved in making the switch].'

Alphabetic:

The alphabetic keyboard is arranged from A-Z. With this logical arrangement of keys the keyboard is meant to be easier for inexperienced typists, however Norman and Fisher (1982) pointed out that results from the available studies disagreed with this theory. In fact it was discovered that for semi-skilled typists the key rates and error corrections were better using the QWERTY while novices seemed to perform equally on either (Michaels, 1971).

b) Pointing Devices

Touch Screens

These enable a user to input information by merely touching a screen at an appropriate point. By combining both the presentation and entry of information in one device, the screen, there are advantages:

- direct eye-hand co-ordination, making them easy and quick to use (Hopkin, 1971; McEwing, 1977),
- display of all valid inputs on the screen, requiring no memorisation of commands and reducing errors (Greenstein and Arnaut, 1988),
- use of a natural pointing gesture, minimising training time (Usher, 1982),

However, there are disadvantages, as Pfauth and Priest (1981) describe:

- initial high cost for the system,
- increased programmer time,
- possible screen glare,
- physical fatigue from reaching to the screen,
- hand blocking of the operator's view of the screen.

In addition, a user must sit within arms reach of the display and, regardless of the screen resolution, target size is determined by the size of the operator's finger, i.e. touch screens are inappropriate for the selection of small objects. A further problem, noted by Shneiderman (1987), was that the software accepts the touch immediately, denying the user the opportunity to verify the selection made.

Light pens

These enable the user to input information by pointing to a spot on the screen with the light pen. They can be used effectively to position a cursor or to select responses. The advantages and disadvantages are very similar to those for a touch screen. However, a light pen is much cheaper than a touch screen, though it has an additional problem with parallax when pointing to objects at the side of the display. This requires such objects to be made larger to counteract the effect of incorrect placement.

Joysticks

Indirect pointing devices, such as a joystick and a mouse, eliminate the problems of hand-fatigue and of obscuring a screen by a hand but have problems associated with (Shneiderman, 1987):

- moving the hand to the device and back again,
- requiring more cognitive processing and hand-eye co-ordination to bring the cursor to the target
- gaining accuracy, resulting in the need for large target objects.

Joysticks are appealing for tracking purposes, i.e. following moving objects on a screen, because of the relatively small displacements required to move the cursor and the ease of directional changes. Oborne (1987) suggests that to aid precision, the joystick should be designed to enable an operator to rest the wrist whilst making movements, and that the pivot point should be positioned under the resting place for the wrist.

Mouse

A mouse is appealing because the hand can rest comfortably on top, the buttons on the mouse are easily pressed, long motions can be rapid and positioning can be fairly precise (Lu, 1984). However, the hand must be moved to the mouse and back, it uses desk space, is difficult for left-handed users, is over sensitive, and some practice is required to develop skill (Shneiderman, 1987; Berry and Broadbent, 1987). In addition, it can move during confirmation, thus there is a difficulty in selecting small objects.

c) Comparison of pointing devices

Human factors such as speed, accuracy, learning time as well as personal preferences, all play a part in how a device is perceived by a user. There have been a few comparative studies.

Card et al (1978) compared four input devices on a text selection task. These devices were: a mouse, a joystick, step keys (four keys, one each for up, down, left, right one line or one character at a time), and text keys (function keys to placer the cursor at the previous or next character, word, line, or paragraph). All devices except the mouse, required the opposite hand to be used to press a confirmation button which was separate from the device itself. The mouse required one of its buttons to be pressed. Total response times, positioning times and error rates were measured. The mouse was superior to all the other devices in all of these categories. Therefore Card et al concluded that the mouse required less mental effort to use and was better at moving and positioning the cursor around the screen.

Karat et al (1986) compared a touch screen, a mouse and a keyboard for target selection, menu selection, and menu selection with typing tasks. When selecting an option with the keyboard, just the letter associated with the choice was required. There was no confirmation needed for the keyboard or touch screen, though the mouse required a mouse button to be used. They discovered that target selection was faster with the touch screen and keyboard rather than the mouse. Menu selection (with and without typing sub-tasks) were faster with the touch screen, followed by the keyboard, and then the mouse. The participants in the study reported that they preferred the touch screen and keyboard to the mouse. Karat et al suggested that since target selection is a practised skill, the pointing action required by the touch screen was more 'natural' and therefore required less cognitive processes than the actions required by the others.

The results from the two studies may not actually be in conflict since Card et al (1978) studied text selection whilst Karat et al (1986) investigated target selection and menu selection. Card et al also used step keys and text keys rather than alphanumeric keys associated with the target. Finally, Card et al required confirmation actions with all devices, though the mouse had an integrated confirmation button. Karat et al only required confirmation on the mouse device. Greenstein and Arnaut (1988) believe that the difference in results from the two studies may in fact reflect this overhead of introducing a confirmation action. The observations of Albert (1982) certainly seem to support this suggestion.

In a different study, Ewing et al (1986) discovered that if there are a few targets on the screen and the cursor can be made to jump from one target to the next, the cursor keys could become the fastest selection device. In addition, MacLean et al (1985) and Berry and Broadbent (1987) both found that if typing and pointing are required in a task, cursor key selection was faster than mouse selection and the method preferred by the users.

However, as Greenstein and Arnaut (1988) warn, it is difficult to draw generalisations from the studies undertaken since:

- most compare only a limited number of devices,
- the devices used vary across the studies,
- confirmation differences exist within and across studies,
- differences in tasks, training and users, result in performance differences with the same device.

Therefore the selection of an input device for a specific application should involve the following considerations:

- characteristics of the task, users, working environment and existing hardware,
- present and future demands of the application,
- the various advantages and disadvantages of the devices in a variety of tasks (see table 1.1),
- user preferences since it is important to provide users with a tool they will use,
- the monetary and development costs involved.

	Touch Screen	Light Pen	Mouse	Joystick
Eye-hand Co-ordination	+	+	0	0
Unobstructed view of display	-	-	+	+
Ability to attend to display	+	0	0	+
Freedom from parallax problems	-	-	+	+
Input resolution capability	-	-	+	+
Flexibility of placement within workplace	-	-	0	+
Minimal space requirements	+	+	-	+
Minimal training requirements	+	0	0	0
Comfort in extended use	-	-	0	+
Suitability for:				
pointing	+	+	+	-
rapid pointing	+	+	0	-
pointing with confirmation	-	0	+	-
alphanumeric data entry	-	-	-	-

Table 1.1: Advantages (+) and disadvantages (-) of a few standard pointing devices (compiled from Greenstein and Arnaut, 1988). Note: 0 = Neutral.

To summarise, the studies seem to conclude that the choice of pointing device, or indeed whether the cursor keys will suffice alone, is actually dependent upon user preferences, the monetary costs, the tasks, and, the environment. Investigations have also been carried out in an attempt to determine the best methods for displaying information. The following section briefly summarises a number of the findings.

d) Displays

With regards to the display, there are a number of factors which may seem obvious but which are not always considered by an interface designer. For example,

- colour blindness (8% of men are affected). Red-Green is the most common,
- certain colour combinations and intensities can cause ghost figures, movement and curvatures,
- high intensities and flickering colours can cause eye strain.

Eye strain has been associated with the level of contrast between the background and the character colours. The greater the difference, the more the eyes suffer. This resulted in a number of countries, for example Sweden, only permitting brown screens with amber writing to be used. Black and white displays were banned.

Technology has advanced and the quality and resolution of screens have greatly improved. Glaring differences between colours can now be avoided by the subtle use of shades. These advancements have enabled designers to benefit from the results of research published in 1980. Radl (1980) and Bauer and Cavonius (1980) discovered that people seemed to prefer negative contrast, i.e. a light background with dark characters, and that with this approach, lower error rates were recorded. The reason for this could be linked to newspapers, books, letters, etc. generally using negative contrast, hence people are familiar with this type of display rather than with the positive contrast approach.

However, users now complain of eyestrain if they have to use software with a predominately bright or very light background. Again the issue of contrast between the characters and the background seems to be the influencing factor. In addition, in the past higher luminance, which exists with negative contrast, made the visual system more sensitive to flicker. This was extremely irritating and was especially noticeable at 50 Hz or on large screens (Hulme, 1984).

Only recently has hardware been developed capable of presenting displays in negative contrast without major problems due to flicker. Microsoft Windows demonstrates this change in policy very clearly. Most of its displays use a white, or off-white, background with darker characters (though most users quickly change white to a grey or a darker colour to reduce the brightness and contrast of the background with the colour of the characters). Other displays have a 'natural' background, for example in the card game 'Patience' the background is green resembling the green baize of the card table. Consequently, with the improved technology now available, an interface designer can choose and select colours that are 'natural' to a user's task, i.e. colours that end-users associate with the task environment.

Colour displays are attractive to users and can lead to rapid recognition and identification of required facilities (Christ, 1975; Robertson, 1980). However, if designers do not standardise the colours throughout their interfaces or use colours inappropriately, colour can actually inhibit performance and confuse users. The danger of misuse is high, hence care and consideration must be taken (Durrett and Trezona, 1982; Shneiderman, 1987).

Another factor, not often considered in interface design but which does seem to influence clarity and reading speed, is the use of capital letters. Tinker (1965) found that upper-case text reads 14% - 20% slower than text which uses both upper and lower case letters. The belief is that shape plays a part in word recognition and capital letters remove this information. Underlining can also effect the way the shape of a word is perceived, making it harder to read. Bruder (1978) found that highly familiar words were effected more by distortion of their shape than less frequently used words. This does seem to indicate that the shape of a word is important, since common words, which people tend to scan over, would be the ones that are effected more by any type of distortion than those words which are unfamiliar and which are read more thoroughly anyway.

How information is presented on a screen, in what colour, size or character form, influences greatly the legibility and the readability of the text. Eye strain, fatigue, headaches and stress have all been associated with badly designed computer displays (Hulme, 1984). The indications are therefore that there is more to the art of presenting information than merely printing text on a screen. For example, a number of important factors that have emerged are:

i) the amount of information to present

Empirical evidence has demonstrated that human performance, i.e. in terms of time and errors, tends to deteriorate with an increase in the display density. The optimum coverage is believed to be around 30% (Dodson and Shields, 1978; Smith and Mosier, 1986; Tullis, 1984; 1988). Consideration should be given to such factors as:

- use of appropriate abbreviations, if they are well known to users,
- avoidance of unnecessary detail,
- use of concise wording,
- use of familiar data formats.

ii) the grouping of information

Most guidelines stress the importance of grouping, (e.g. Danchak, 1976) although little empirical evidence exists directly linked to the grouping of information. Common agreement exists stating that users should be able to assume that elements within a group are all somehow related to each other semantically. Elements can be grouped by:

- spatial proximity,
- colour,
- boundaries,
- highlighting.

iii) the highlighting of information

Among the most commonly used highlighting techniques are:

- reverse video
- colour
- underlining
- flashing.

However, although highlighting is an effective technique, two important factors must be remembered conecrning its use. Firstly, it should be applied conservatively, and secondly, the items to be highlighted must be selected carefully. Since highlighting attracts users attention, highlighting too many items or the wrong items will distract a user from important pieces of information on display.

iv) the placement and sequence of information

Every screen should be laid out in a manner which allows a user to find any information easily. One of the best methods is to adopt a consistent format for all the screens in an application. This allows a user to develop expectancies regarding the positioning of information on the screen, hence making the application easier to learn. This has been demonstrated empirically by, for example, Tullis, 1981, Teitelbaum and Granda, 1983. The optimum sequence of presenting data on a screen is determined by a number of factors, such as:

- sequence of use
- importance
- frequency of use
- alphabetic or chronological order

v) the presentation of text

There are many guidelines regarding the presentation of text (e.g. Galitz, 1985; Smith and Mosier, 1986; Tullis, 1988). Some of the issues discussed are:

- the case of the letters (as explained above),
- the justification and spacing between words,
- the indentation used.

Consequently, these previous studies, as well as research findings from the fields of visual perception and cognitive psychology, have resulted in a number of useful guidelines being written for interface designers (for example, Smith and Mosier, 1984; 1986; Shneiderman, 1987; Heckel, 1984; Tullis, 1983; Van Ness, 1986). Since well presented information improves both readability and recognition, and therefore the reader's understanding and comprehension of the text, many of these issues need to be considered and reviewed carefully. Five major points, for both data display and data entry that have emerged from these guidelines are:

a) constistency of data entry and display, e.g. format, positioning, actions, messages, terminology, abbreviations,

b) efficient information assimulation by a user, e.g. terminology, format and sequence that are natural to a user, task and environment,

c) minimal memory load and minimal input actions required by a user, e.g. non-redundacny, meaningful messages, task completion with a few commands, use of selection rather than free-format entry, reduce window and task switching,

d) supportiveness of the system, e.g. feedback, reversal of actions, help, confirmation messages, error handling.

Shneiderman (1987) concludes, 'these underlying principles must be interpreted, refined and extended for each environment. The principles presented focus on increasing the productivity of users by providing simplified data entry procedures, comprehensible displays and rapid informative feedback that increase feelings of competence, mastery, and control over the system'. Consequently, it is crucial that developers consider carefully all of these criteria when designing software otherwise the end-user acceptability of the finished system will be seriously undermined.

References

Albert, A.E., 1982, 'The effect of graphic input devices on performance in a cursor positioning task', in Proc. Human Factors Soc., 26th Annual Meeting, 1982, p54-58

Alden, D.G., Daniels, R.W. and Kanarick, A.F., 1976, 'Keyboard design and operation: a review of the major issues', in Human Factors, Vol.14, p275-293

Bauer, D. and Cavonius, C.R., 1980, 'Improving the legibility of visual display units through contrast reversal', in Ergonomic aspects of visual display terminals, (eds) Grandjean, E., Vigliani, E., London:Taylor and Francis, p137-143

Berry, D.C. and Broadbent, D.E., 1987, 'Expert systems and the man-machine interface. Part two: The user interface', in Expert System, Vol.4, No.1, p18-28

Bruder, G.A., 1978, 'Role of visual familiarity in the word superiority effects obtained with the simultaneous-matching task', in Journal of Experimental Psychology: Human perception and performance, Vol.4, p88-100

Card, S.K., English, W.K. and Burr, B.J., 1978, 'Evaluation of the mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT', in Ergonomics, Vol.21, p601-613

Christ, R.E., 1975, 'Review and analysis of colour coding research for visual displays', in Human Factors, Vol.17, No.6, p542-570

Danchak, M.M., 1976, 'CRT displays for power plants', Instrumentation Tech. Vol.23, No.10, p26-36

Dodson, D.W. and Shields, N.E., 1978, 'Development of user guidelines for ECAS display design', Vol.1, report No. NASA-CR-150877, Essex Corp. Huntsville

Dunn, A.G., 1971, 'Engineering the keyboard from the human factors viewpoint', in Computers and Automation, Feb. 1971, p32-33

Durrett, J. and Trezona, J., 1982, 'How to use colour displays effectively', in Byte, April, 1982, p50-53

Ewing, J., Mehrabanzad, S., Sheck, S., Ostroff, D. and Shneiderman, B., 1986, 'An experimental comparison of a mouse and arrow-jump keys for an interactive encyclopedia', in Int. J. of Man-Machine Studies, Vol.23, p29-45

Galitz, W.O., 1985, Handbook of screen format design (2nd edition), Wellesley Hills: QED Info. Sci.

Greenstein, J. and Arnaut, L., 1988, 'Input devices', in (ed) Helander, M., Handbook of Human-computer Interaction, North-Holland, p495-516

Heckel, P., 1984, The elements of friendly software design, New York: Warner Books

Hopkin, V.D., 1971, 'The evaulation of touch displays for air traffic control tasks', in IEF Conf. on Displays: Publ. No. 80

Hulme, C., 1984, 'Reading: extracting information from printed and electronically presented text', in Fundamentals of human-computer interaction, (ed) Monk, A., London:Academic Press, p35-46

Karat, J., McDonald, J.E. and Anderson, M., 1986, 'A comparison of menu selection techniques: touch panel, mouse and keybaord', Int. J. Man-Machine St., Vol.25, p73-88

Lu, C., 1984, 'Computer pointing devices: Living with mice', in High Technology, Jan. 1984, p61-65

MacLean, A., Barnard, P.J. and Wilson, M.D., 1985, 'Evaluating the human interface of a data entry system: user choice and performance measures yield different tradeoff functions', in People and Computers: Designing the interface, (ed) Johnson, Cambridge Press Publ., p172-186

Martin, A., 1972, 'A new keyboard layout', in Applied Ergonomics, Vol.3, p48-51

McEwing, R.W., 1977, 'Touch displays in industrial computer systems', in Displays for Man-Machine Systems, London:IEE, p79-81

Michaels, S.E., 1971, 'QWERTY versus alphabetic keyboards as a function of typing skill', in Human Factors, Vol.13, p419-426

Norman, D.A. and Fisher, D., 1982, 'Why alphabetic keyboards are not easy to use: Keyboard layout doesn't matter much', in Human Factors, Vol.24, p509-519

Oborne, D.J., 1987, 'Ergonomics and new technologies', in Information technology and people, (ed) Blackler, F., Oborne, D., Brit. Psychological Society Publ., p65-85

Pfauth, M. and Priest, J., 1981, 'Person-computer interface using touch screen devices', in Proc. of 25th Annual Meeting of Human Factors Soc., Baltimore: HFS,p500-504

Radl, G.W., 1980, 'Experimental investigations for optimal presentation-mode and colours of symbols on the CRT-screen', in Ergonomic aspects of visual display terminals, (eds) Grandjean, E., Vigliani, E., London:Taylor and Francis, p127-137

Robertson, P.J., 1980, 'A guide to using colour on alphanumeric displays', IBM Tech. Report G320-6296-0, White Plain, NY

Shneiderman, B., 1987, Designing the user interface: strategies for effective human-computer interaction, Wokingham: Addison-Wesley Publ.

Smith, S.L. and Mosier, J.N., 1984, 'Design guidelines for the user interface for computer- based information systems', The MITRE Corp., Bedford, MA 01730, Electronic systems division

Smith, S.L. and Mosier, J.N., 1986, 'Guidelines for designing user interface software', Report ESD-TR-86-278, The MITRE Corp, Bedford

Teitelbaum, R.C. and Granda, R.E., 1983, 'The effects of positional constancy on searching menus for information', Proc. CHI '83 Conf. Human Factors in Computing Systems, New York: Assoc. for Computing Machinery, p150-153

Tinker, M.A., 1965, Bases for effective reading, University of Minnesota Press

Tullis, T.S., 1981, 'An evaluation of alphanumeric, graphic, and colour information displays', Human Factors, Vol.23, No.5, p541-550

Tullis, T.S., 1983, 'The formatting of alphanumeric displays: a review and analysis', Human Factors, Vol.25, No.6, p557-582

Tullis, T.S., 1984, Predicting the usability of alphanumeric displays, Doctoral dissertation, Uni. Michigan, USA

Tullis, T.S., 1988, 'Screen design', In (ed) Helander, M., Handbook of Human-computer Ineration, North-Holland

Usher, D.M., 1982, 'A touch sensitive VDU compared with a computer-aided keypad for controlling power generating plant', a paper presented at IEE Conf. on Man-Machine Systems

Van Ness, F.L., 1986, 'Space, colour and typography on visual display terminals', Behaviour and Info. Tech., Vol.5, p99-118

Appendix C (I)

Questionnaire production trial

The first investigation undertaken was intended both to determine whether the word-processed files produced by secretaries were accurate representations of the submitted hand-written questionnaires and to identify the common layout technique used by secretaries.

The questionnaires had to be produced in ASCII format ready for the IDDA generator to read. It was thought that if secretaries could complete this task, their typing skills and knowledge of word-processors could be utilised. The initial investigation involved six secretaries. Each was given the original questionnaires for the Patello-femoral clinical trial and was asked to type the separate assessment questionnaires into an ASCII text file using a word-processor. They all had previous knowledge of at least one word-processing package prior to the trial.

The results of the evaluation revealed that the secretaries were quite happy with and competent at producing the required ASCII text file, but that they tended to adopt different layout techniques. These differences could cause problems later when the files were being read by the IDDA generator. For example, two secretaries added question numbers, four placed a blank line between the question and the answer list, three placed extra lines between questions, and, four used tabs to align the questions and answer lists, whilst the other two just used spaces. In addition, some word-processors placed extra blank lines at the end of a page when the file was converted to ASCII text.

The IDDA generator is required to separate the questions from the questionnaire, one at a time, whilst attempting to determine the question type, i.e. comment, date, numeric etc.. Therefore the layout is quite important. Consequently, a tool had to be developed to attempt to accommodate these differences. It had to:

a) check to ensure that there were no question numbers in the file.

   The IDDA tools generate question numbers themselves and these numbers are used to reference the individual data items and to select questions. Consequently, confusion would occur if the question numbers, added by a secretary, differed from the ones generated by the tools. There had to be consistency. Automatic generation of numbers requires less typing, assists in question alignment, ensures accuracy, and takes into account the inclusion of comments. Therefore automatic question number generation was considered to be the best approach to adopt.

b) allow comments to be specified.

   From this study, and from the review of a number of old questionnaires, it was evident that comments must be able to be classified, i.e. text to which no answer was expected. A comment, it seems, can be placed either before, or after, the question to which it is linked. If a question does have a comment, both must appear on the same data entry screen for the comment to be of any use to the operator.

c) check the total length of the question and answer box, including any attached comment or answer list.

The maximum length of any question is 15 lines to ensure that there is still room for the title and the record identification to be displayed. Therefore there needs to be a check on the size of a question, its answer, any attached comment and/or answer list.

d) determine the beginning of a new question or comment.

It was evident that the secretaries used the same margin size for all questions, i.e. all questions were aligned. In addition, they would indent the answer list attached to a question and would insert a blank line between questions. Sometimes blank lines were used between a question and its answer list and/or between each line in the answer list. For example,

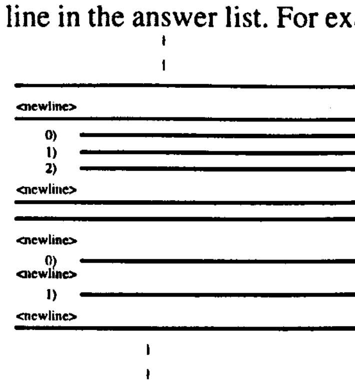


Diagram 1: Two question layouts that were identified

Consequently, the test for a question (or comment) could be determined by the margin size and the position of blank lines. For example, if there was a text line with a margin of size x, then a blank line, followed by another text line with the same margin size, these are two different questions (or a question and a comment) (see diagram 2, part a). If no blank line exists, the lines are part of the same question (see diagram2, part b).

If there was a text line with margin of size x, then a blank line, followed by another text line with a margin of size x+y, and the second text line began with a number, these must form a question and its answer list (see diagram 2, part c). A similar conclusion can be drawn if there was no blank line in the above example (see diagram2, part d).

If the second text line began with a '(', it is likely to be an extended explanation of the question (see diagram 2, part e).

If, however, the second line had a margin of size x+y and did not start with a '(' or a number and there was a blank line between the two, it would be the start of another question or a comment with a mistyped margin, e.g. one extra space character than was needed (see part f).

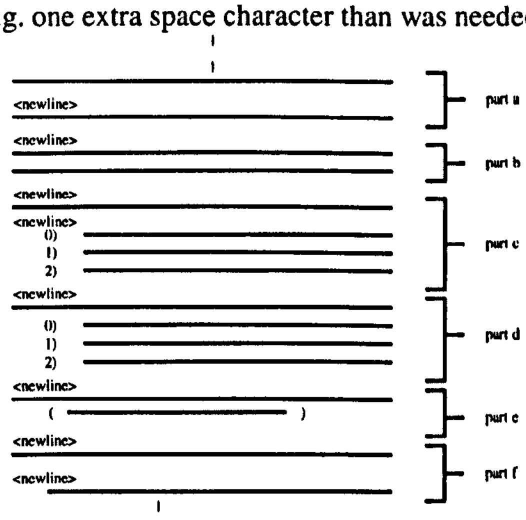


Diagram 2: Possible question and answer layouts

Consequently, from a knowledge of the different layout styles, the tools can be designed to separate the various questions and comments from a text file.

e) remove any extra lines between questions but leave one blank line between a question and its answer list, if a blank line was already there.

This was to ensure that as many questions as possible could fit on a screen, without compromising any reasonable layout defined by the user.

f) convert all tabs to spaces.

If a consistent number of spaces is used throughout the conversion then the original alignment with tabs should not be affected by the change. If the tabs were merely removed and no spaces substituted then the desired layout would disappear. As tabs are control characters, leaving them in would cause problems later. Consequently, the tools convert the tabs to spaces.

g) attempt to determine the answer type for a question.

From the literature (e.g. Pocock, 1985) and from reviewing previous clinical questionnaires, it is clear that the advice 'virtually all questions should be constrained so that the answer can be given in numerical format' (Pocock, 1985) is followed closely. Even qualitative alternatives are given associated numbers (e.g. male = 1, female = 2). This is because more open-style answers are harder to incorporate in statistical analysis and would therefore be kept merely as background data which may never be used.

Therefore, if an answer list is present with a question, the answer type is highly likely to be numeric. Moreover, the number on the first line after the question can be extracted as can the number of the last line before the next question. These values can be used as the minimum and maximum values of the answer range. Furthermore, numbers which are used in answer lists are whole numbers, thus there would be no need for any decimal places.

If, on the other hand, there is no answer list and the question contains the word 'date' once (in any mixture of cases), the answer type is likely to be a date field. Although this is not certain, for example, 'what is the age of a patient at the date of assessment?', another guideline in designing clinical forms advises that dates should be recorded rather than expecting the investigator to calculate the time interval. Therefore, there is a good chance that if 'date' is met in the question then the answer itself will be a date.

If neither of the above is true and the question has 'Y/N', 'YES', 'T/F' or 'TRUE' (again in any case), the answer type is likely to be logical. If none of the above apply, the answer type is likely to be alphanumeric.

The tools are not capable of inferring the correct answer type every time. Therefore, a computer-generated suggestion would not be consistent, e.g. neither always correct nor incorrect. If it is displayed to the user, it could influence the choice of a naive operator, i.e. the 'computer is always right' syndrome. Thus, it could be a dangerous facility. There is no gain in entry speed through displaying the suggestion as a default value, as the user must still check the answer type. Consequently, the computer-generated suggestion is used as a hidden guide, i.e. the suggested answer type is not displayed but is used to check the entry of the operator and, if they differ, a message appears prompting users to confirm their selection. In this manner, users will be answering all of the questions in a consistent way, thus increasing the speed of operation and the understanding of users. It also assists users in perceiving that they are in control of the task.

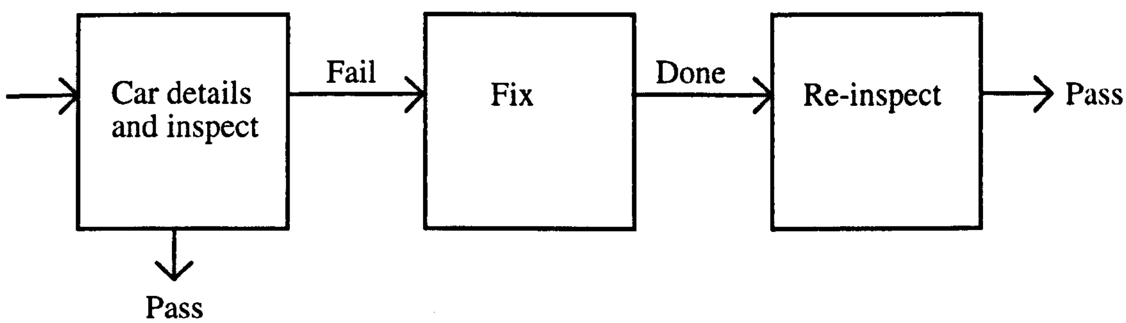
Consequently, an automatic approach to adapting and checking the entered questionnaires was used to prevent requesting the secretaries to follow one predefined format, which is highly likely to be alien to them in some respect. They can thus use whichever layout technique is most natural to them, so long as the two basic rules regarding question length (max. 15 lines) and the absence of question numbers are followed. It is believed that this approach would reduce errors and increase typing speed since the secretaries will be more relaxed and at ease whilst completing their task.

Appendix C (II)

Ministry Of Transport (MOT) trial

The first trial reviewed the whole automatic generation process to determine whether the tools could construct the required IDDA end-system from the information provided. The test selected was a straight-forward MOT system, based around the form completed by garages when a car is brought in for an MOT (see page C-6).

It entailed 3 stages:



It was envisaged that such a system could assist in establishing the life expectancy of individual parts and help in determining the problems associated with particular types of car. If a given car was deemed dangerous, this would enable it to be called in early for an inspection. The information could also be used to estimate more accurately, both in terms of time and money, the maintenance costs for different makes of cars.

For this initial evaluation, the following questions were being examined:

- is the proposed methodology appropriate to build an IDDA end-system,
- did the design and building of the end-system follow a logical path,
- were the interfaces pleasant and easy to understand,
- were the tools easy to use,
- did the automatic generator produce a workable end-system,
- what other facilities would be required in the tools and/or in the end-system to further assist the operator.

Both my supervisor and myself carried out this MOT trial as the objectives did not include reviewing whether the tools were appropriate for naive computer users. The comments emerging from the evaluations are an amalgamation of our views.

Points to note from the trial

The IDDA system was constructed successfully by both evaluators. The resultant IDDA end-systems were 'reasonable' with respect to the tasks for which they had been designed and constructed. In other words, these systems were accurate representations of the coupling of the specifications entered into the tools by the user and the tool's own internal computing knowledge.

The pathway leading the user through the design and building phases did seem to be both logical and easy to follow. In addition, the tools themselves were deemed to be relatively easy to use. However, the views were that the interfaces required more colour and that certain questions needed to be rephrased into a shortened form. Moreover, it was felt that the facilities offered needed to be enhanced in the following manner:

Building the end-system

- the ability to go back to re-specify characteristics of the current question being defined,
- the ability, during the entry of the information, to quit and save at any point,
- the ability to resume entry from the interrupted point,
- the ability to quit and re-start building the end-system from scratch,
- the ability to use alphanumeric identifiers,
- the provision of a more detailed HELP facility,
- the ability for lists and tables to be defined,
- the ability to specify that no user defined HELP was to be linked into the IDDA end-system.

End-system

- function keys to: page forward, escape and save, exit and discard, and, enter null values.

There was also the necessity to hide from the user, the messages produced during the compilation of the generated programs.

These evaluations seemed to indicate that the pathway devised for designing and building an IDDA end-system was appropriate and that it could be applicable to certain applications outside medicine. Moreover, the code that was automatically produced by the tools was 'correct' in terms of both utilising the information entered by the user and compiling successfully to produce a reasonable working end-system. However, the evaluations also uncovered both the need for additional facilities to be made available to users, especially naive computer users, and the opinion that the general layout of the interfaces for the tools required improving. These alterations would be carried out prior any subsequent trials, which could then review the changes made.

Vehicle Reg. Mark ______________________ VIN or Chassis no. ______________________

Make & Model ______________________ Colour __________

A Testable Item	Testers Manual Ref. (see over)	Pass	Fail	Reason for Failure and Remarks
Section I — Lighting Equipment				
Front Lamps	I/1			
Rear Lamps	I/1			
Headlamps	I/2			
Headlamp Aim	I/6			
Stop Lamps	I/3			
Rear Reflectors	I/4			
Direction Indicators	I/5			
Section II — Steering & Suspension				
Steering Controls	II/1			
Steering Mechanism	II/2			
Power Steering	II/3			
Transmission Shafts	II/2,4,9			
Stub Axle Assemblies	II/5			
Wheel Bearings	II/4			
Suspension	II/5,6,7,8,9			
Shock Absorbers	II/10			
Section III — Brakes				
Service Brake Condition	III/3,4			
Parking Brake Condition	III/1,2			
Service Brake Performance	III/5,6,7,8			
Parking Brake Performance	III/5,6,7,8			
Service Brake Balance	III/5,6,7,8			
Section IV — Tyres & Wheels				
Tyre Type	IV/1			
Tyre Condition	IV/1			
Roadwheels	IV/2			
Section V — Seat Belts				
Security of Mountings	V/1			
Condition	V/1			
Operation	V/1			
Section VI — General Items				
Windscreen Washers	VI/1			
Windscreen Wipers	VI/2			
Horn	VI/4			
Exhaust System	VI/3			
Silencer	VI/3			
Vehicle Structure	VI/5			

B **Test Result** 1 ☐ PASS Test certificate issued. No: ______________________

2 ☐ FAIL see below

C **Notice of Refusal of a Test Certificate** (see notes overleaf)

1 ☐ For the reasons shown in the above Inspection Report

2 ☐ Because the inspection could not be completed, for the following reasons:

D **Warning** In my opinion the vehicle is DANGEROUS to drive because of the following defects:

E

Signed ______________________ Date ______________________

(Tester/Inspector)

NAME (BLOCK CAPITALS) ______________________ Testing Station no. ______________________

YOU ARE ADVISED TO KEEP THIS FORM

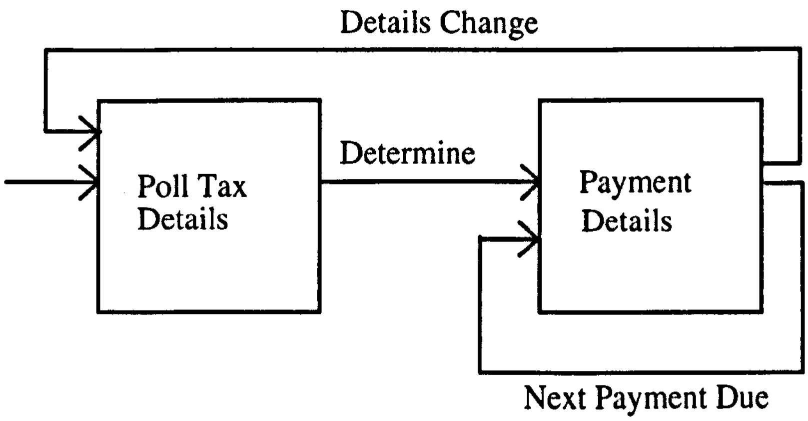
Appendix C (III)

Poll Tax trial

This evaluation was carried out to test and review the alterations undertaken after the MOT evaluation.

The task was to build a simple Poll Tax recording system. This was chosen as it was another application area outside the medical domain and, at the time, it was a topic for which there existed plenty of information.

Rather than a 3 stage system, as with the MOT trial, this evaluation required the building of a 2 stage assessment process (see pages C-9 through to C-13 for details of the Poll Tax questionnaires and the answer details for the questions in the questionnaires). In this sytem, the first questionnaire is used to collect a client's details and assess a Poll Tax bill for a client. The second questionnaire monitors payment details and any fines that may have accrued due to a client's refusal to pay. A client's situation may change, for example moving house, and therefore, there needs to be the possibility to change the details of a client. A schematic diagram of the information cycle can be seen below:



Once more, the operators of the tools were my supervisor and myself.

Points to note

Both the Poll Tax systems were built successfully. All the facilities which had been added after the MOT trial worked correctly.

Moreover, the requirement for the user to summarise the question from the questionnaire, enabling a variable name to be generated, was removed. This was replaced by the use of the tools to automatically generate the variable name from the information already given. This resulted in less information requiring to be entered by the user and thus time was saved, typing errors reduced, and user frustration decreased.

Queries:

A number of queries arose from this evaluation:

a) The ability to calculate results from cells of a table and/or list and have the answer as a response to a further question

b) The ability to have sections, i.e. group questions in the questionnaire together, enabling the whole section to act in a particular way, i.e. be skipped if a previous answer means that the section is irrelevant

c) The ability to have an XOR facility

d) The ability to view the previous questions in the questionnaire on second monitor during defining the condition linked to the current question

Responses:

The following are the responses to these queries:

a) The ability to automatically calculate results of previous questions within the questionnaire and have the answer as the response to another question, could be added later, if deemed necessary. The implementation of this facility is quite straightforward and the tools could be programmed to check that the selected questions were either date or numeric. Before adding this, the intended user group would need to demonstrate that such a facility was required.

b) Again, the requirement for sections must be demonstrated by the intended user group, especially as individual questions can already be programmed to react in the required manner.

c) Generally only mathematical and/or scientific people know XOR, therefore an implementation of XOR is likely to confuse people. XOR can be represented by the use of ORs and ANDs, if it is required. The majority of people think in an inclusive manner and consequently, this should be the norm.

d) To build up the questionnaire questions on the second monitor, as they are defined, would require the second monitor to be frequently refreshed. Moreover, as the second monitor is used for displaying HELP, the list would disappear everytime the user requested HELP. It would then require re-generation once the user had quit the HELP facility. This would slow the tools down. Instead, since the users of the tools are highly likely to have the paper version of the questionnaires in front of them during the definition stage, it is felt that they are more likely to refer to the paper copy than to the second screen, if they required assistance. Hence it was decided that this facility would only be added if, after subsequent trials, a significant number of users indicated a need for this ability.

Consequently, the queries which emerged from this evaluation were much broader in scope than those which evolved from the MOT trial. However, for answers to these types of question, investigations in a more realistic environment with users from the proposed user group must be completed to determine whether these features are in fact required by the intended operators.

Poll Tax Trial

Initial Questionnaire

Reference Number

Date

Are you :- < 18
Mentally Impaired
Hospital / Care home patient
Prisoner
Member of Religious Community
Foreign Diplomat
Visiting Army personnel / Non-British dependants
Volunteer Care Workers
Resident of Some Crown Building

Are you a Full-time student

Name of institution

Address of institution

Title of Course

Course start date

Course finish date

Rebates

Are you on income support / wage = income support

Are you married / cohabiting with opposite sex

Title of partner

First name of partner

Surname of partner

What is your annual wage

What is your annual tax

What is your N.I.

What are your annual benefits
(Pensions, Child Benefits, Disability Allowance,
Invalidity Benefit, Statutory Sick Pay)

What is your joint annual wages

What is your joint annual tax

What is your joint N.I.

What are your joint annual benefits
(Pensions, Child Benefits, Disability Allowance,
Invalidity Benefit, Statutory Sick Pay)

Do you have < £7000 in savings

Do you have > £16000 in savings

Amount of savings

Do you jointly have < £7000 in savings

Do you jointly have > £16000 in savings

Amount of joint savings

Enter income support level for situation

Transitional Relief

Are you a private tenant paying inclusive rates

Have you moved since 1st April 1990

Were you 18 on or after 1st April 1990

Are there 2 or more poll tax payers in the household

How many

Actual poll tax for area

Assumed poll tax for area

Are you a pensioner (> 60 female, > 65 male)

Are you disabled - Attendance Allowance Mobility Allowance
Supplement, Invalidity Pension, Severe
Disablement Allowance Registered Blind
Disabled

Have you previously paid rates

Assumed rates for property 1989-90

Enter year number - 1990-91 = 1
1991-92 = 2
1992-93 = 3
1993- = 4

Poll Tax Trial

Initial Assessment

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is question asked?
1	Ref. Number	N	0	999999	0	-----	0	Always
2	Date	D	-----	-----	-----	-----	<	Always
3	Are you : - <18	L	-----	-----	-----	-----	N	Always
4	Full-time student	L	-----	-----	-----	-----	N	Always
5	Name institution	A	-----	-----	-----	30	(blank)	4 = 'Y'
6	Address instit.	A	-----	-----	-----	80	(blank)	4 = 'Y'
7	Title course	A	-----	-----	-----	30	(blank)	4 = 'Y'
8	Course start date	D	-----	-----	-----	-----	<	4 = 'Y'
9	Course fin. date	D	-----	-----	-----	-----	<	4 = 'Y'
10	**Rebates**	**C**	-----	-----	-----	-----	-----	**<Next question>**
10	Income supp.	L	-----	-----	-----	-----	N	Always
11	Married	L	-----	-----	-----	-----	N	Always
12	Title partner	A	-----	-----	-----	10	(blank)	11 = 'Y'
13	Surname partner	A	-----	-----	-----	20	(blank)	11 = 'Y'
14	Forename	A	-----	-----	-----	20	(blank)	11 = 'Y'
15	Annual wage	N	2	999999	0	-----	0	11 = 'N'
16	Annual tax	N	2	999999	0	-----	0	11 = 'N'
17	N.I.	N	2	999999	0	-----	0	11 = 'N'
18	Annual benefits	N	2	999999	0	-----	0	11 = 'N'
19	Joint wage	N	2	999999	0	-----	0	11 = 'Y'
20	Joint tax	N	2	999999	0	-----	0	11 = 'Y'
21	Joint N.I.	N	2	999999	0	-----	0	11 = 'Y'
22	Joint benefits	N	2	999999	0	-----	0	11 = 'Y'
23	< 7000 savings	L	-----	-----	-----	-----	N	11 = 'N'
24	> 16000 savings	L	-----	-----	-----	-----	N	11 = 'N'
25	Amount savings	N	2	999999	0	-----	0	11 = 'N' & 23 = 'N' & 24 = 'N'
26	jointly < 7000	L	-----	-----	-----	-----	N	11 = 'Y'
27	jointly < 16000	L	-----	-----	-----	-----	N	11 = 'Y'
28	joint amount	N	2	999999	0	-----	0	11 = 'Y' & 26 = 'N' & 27 = 'N'
29	Inco. supp. lev.	N	0	10	0	-----	0	Always
30	**Trans. Relief**	**C**	-----	-----	-----	-----	-----	**<Next question>**
30	Private tenant	L	-----	-----	-----	-----	Y	Always
31	Moved 1/5/90	L	-----	-----	-----	-----	N	Always
32	18 after 1/5/90	L	-----	-----	-----	-----	N	Always
33	> 2 p.t. payers	L	-----	-----	-----	-----	Y	Always
34	How many	N	0	99	1	-----	2	Always
35	Actual p.t.	N	2	999999	0	-----	0	Always
36	Assumed p.t.	N	2	999999	0	-----	0	Always
37	Pensioner	L	-----	-----	-----	-----	N	Always
38	Disabled	L	-----	-----	-----	-----	N	Always
39	Prev. paid rates	L	-----	-----	-----	-----	Y	37 = 'Y' or 38 = 'Y'
40	Assumed rates	N	2	999999	0	-----	0	37 = 'N' & 38 = 'N'
41	Year number	N	0	4	0	-----	0	Always

Poll Tax Trial

Follow-up Questionnaire

Reference Number

Date

Date of next payment

Payment type -

Payment interval -

Refusal to pay

Fine amount - £50-200

Refusal to pay

Amount due (Balance + cost + fine)

Refusal to pay

Issued Distraint Order etc.

Poll Tax Trial

Payment Assessment

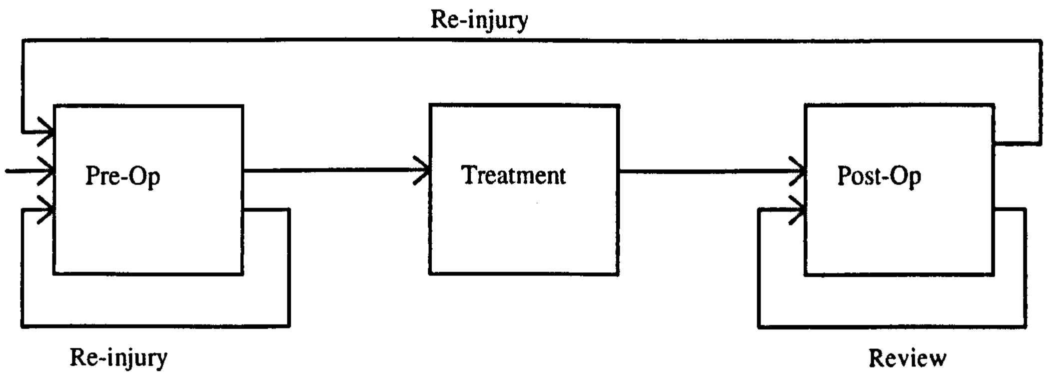
Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Number of entries	Common Answer	When ?
1	Ref. Number	N	0	999999	0	-----	0	Always
2	Date	D	-----	-----	-----	-----	<	Always
3	Date next paym.	D	-----	-----	-----	-----	<	Always
4	Payment Type	S	-----	-----	-----	5 (L)	N	Always
5	Payment interval	S	-----	-----	-----	4 (L)	N	Always
6	Refusal to pay	L	-----	-----	-----	----	N	Always
7	Fine amount	N	2	200	50	-----	50	6 = 'Y'
8	Refusal to pay	L	-----	-----	-----	-----	N	Always
9	Amount due	N	2	999999	0	-----	0	8 = 'Y'
10	Refusal to pay	L	-----	-----	-----	-----	N	Always
11	Issued Dist.Order	L	-----	-----	-----	-----	<	10 = 'Y'

Appendix C (IV)

Patella-femoral joint trial

This investigation involved a medical application and was devised to compare the views and results gained from a computer literate user with no medical knowledge against a naive computer user with a medical background but not in the specialist field. Neither suject was knowledgeable of the IDDA system nor had been given any help documentation prior to the trial. Once again, both the functionality and the interfaces of the tools were being reviewed with respect to usability, user acceptability and applicability to the task.

A patella-femoral assessment procedure was used for the trial. This is an assessment for patients who repeatedly report having 'a pain behind the kneecap'. Mr Harding, a senior orthopaedic consultant, rewrote an existing patella-femoral questionnaire into three assessments (see pages xliv - l). He also produced four sections of textual help clarifying the precise procedures to be undertaken by the medical personnel when examining patients. The assessment process is as follows:



The Pre-Operative assessment consisted of 30 questions, the Treatment assessment of 9 questions, and the Post-Operative assessment of 28 questions. These questionnaires and the help text were typed in via a word-processor and the appropriate ASCII files produced prior to the start of this evaluation.

Points to note

For this trial, two users were selected - a physiotherapist and a researcher in computing. Both users successfully built the required IDDA system. It took the user with no computer experience (the physiotherapist), 43 minutes to completely construct the end-system and the user with computer experience 36 minutes. Neither had been given any documentation to fully explain the system nor any help documentation prior to or during the trial. However, they both had access to the on-line help facility. Even so, overall both found the tools easy to use and the instructions, diagrams and error messages, very easy to follow and understand.

One disparity between the two trialists was the length of time it took them to feel confident in using the tools. As expected, the user with computing experience was more confident earlier in the trial than the user with no prior computing experience. However, even the naive computer user recorded that, by halfway through the trial, she was confident and comfortable using the tools.

It was interesting to note that the person with computing experience referenced the help on occasions whilst the person without computing experience did not reference the help at all. One reason could be that the lack of computing experience meant that there was not the tendency to look to the computer for help when a problem occurred, i.e. look for a help key or button. Other factors influencing this lack of awareness could have been that the screens 'appeared cluttered' to this user (though not to the computer experienced operator) and that this user admitted to not having read all the information on the various screens. This meant that she might not have seen that a help key was available and, consequently, never knew she had the ability to reference help. Nevertheless, she still managed to successfully build the required IDDA system with little difficulty.

It was noticable that the user with no prior computing experience found the words 'default value' confusing. The objective of the 'default value' question is to try to acquire, for the question being defined, the value that is most commonly used by medical personnel during a patient examination. This value would be displayed by the operational IDDA end-system as the initial data value for the question. Therefore the operator of the IDDA system would merely need to use the return key to accept this value rather than having to type in the response. This approach reduces the amount of data needing to be entered and, consequently, reduces the chances of typing errors occurring. If the value is different from the displayed value then obviously it can be changed in the normal manner. However, if the most appropriate value has been selected during this construction phase, the user of the IDDA end-system will receive the benefits of this approach. Consequently, to try to alleviate the confusion, the 'default value' question in the tools has now been changed to a question requesting the 'most common value'. Subsequent trials will determine whether this alteration has indeed solved the problem.

In addition, the user with no computing experience found the act of confirming certain entries annoying and time-consuming. The tools try to determine the type of each question in the ASCII questionnaire files, i.e. whether the response to the question in the questionnaire is a number, character, etc. If the selection entered by the user of the tools differs from this, the tools ask for confirmation. Both of the trialists were happy with this 'response type checking'. However, they were unhappy that when they overrode the tools and went on to define the other characteristics of the response, i.e. the number of decimal places, maximum, minimum values, etc., the tools also continued to ask for confirmation after each of these entries (again, this is because the information the user has entered does not match the answers deduced by the tools). This process has now been changed so that when the user overrides the tools at the initial stage of defining the question's response type, the tools will not ask for any further confirmation of subsequent user entries. If, however, the types are the same, i.e. that deduced by the tools and that entered by a user, then, if the response characteristics which follow differ, the system will still ask for confirmation, i.e. confirmation of the entry will be required if the maximum value entered by a user differs from the maximum value deduced by the tools from the questionnaire file. This change should speed up the question definition stage and reduce the annoyance caused to a user. It would also enable the continuation of a provision for user assistance as well as for valuable error checking.

Both trialists indicated their displeasure at the fact that the ESC key returned them to the initial question definition stage, i.e. defining the question type, rather than just back one question. Though this may be valid, it may be a consequence of the previous problem explained above. In the worst case, when the question is numeric, there are 5 responses to enter and this is assuming that the user is on the last entry for that question. Therefore in the majority of cases, the number of responses requiring re-entering will be much less. This issue will be reviewed in subsequent trials to determine whether it is still a factor after the above changes have been made.

The final suggestion from one of these trialists was to place the function keys at the bottom of the screen rather than at the top. This was previously tried by Ashton-Tate (the producers of dBASE). They discovered that the majority of their users preferred the available keys to be listed at the top of the screen. Therefore, they have subsequently changed their software products back to display the labels in this position. In this system, the tools present the question taken from the questionnaire in the top part of the screen. Consequently, the user will be looking at this section of the screen fairly frequently. Therefore the function keys should be easy to locate. As only one trialist has made this point up to now, it will be reviewed during subsequent evaluations to determine if it does in fact cause problems.

The results and comments from this trial do seem to indicate that the lack of computing knowledge does not ultimately affect the user's ability to build the required IDDA system. The only aspect that seems to have been adversely affected was the time, 43 minutes as opposed to 36 minutes. This extra 7 minutes is unlikely to be a critical factor in the context of the task, i.e. in a construction task. Consequently, from this trial the indications are that computing experience is not required to follow the methodology proposed and to successfully use the tools to construct the IDDA end-system. Similarly, the differences in the amount of medical knowledge do not seem to have caused many difficulties. A further trial reviewing the opinions of naive computer users, who are also unaware of the medical field, would be interesting. However, from these initial results, the tools themselves do not seem to require the user to have any knowledge of the specialist domain nor to have had significant computing experience.

Patella-femoral Joint Assessment Chart

Name.. Diagnosis (Pathology) ....
Address..
....
.... Operation ....
Age........ .... Sex..........................
Occupation.. Side ....
Operation date / / .

	Pre-op / /	Post-op / /	Post-op / /	Post-op / /	Post-op / /
Symptoms - Pain					
Instability					
Locking					
Swelling					
Never = 4	Rarely = 3	Occasionally = 2	Constantly = 1		
Function - Stairs					
Sitting					
Sports					
Normal = 4	Difficult = 3	With aid = 2	Unable = 1		
Signs - Crepitus					
Apprehension					
Tenderness					
⇑ Q angle					
Yes = 1	No = 0				
R.O.M. - Flexion					
Extension					
Strength - quadriceps					
M. R. C. 0 - 5					
X-Rays - Wiberg Type					
PT/PL ratio					
O.A - P.F.J.					
O.A - Knee					
Patient assessment					
Normal = 4	Better = 3	Same = 2	Worse = 1		
Post-op complications (specify)					

Figure C (IV): The original Patella-femoral Joint Assessment Chart

Patella-femoral Trial

Initial Assessment

Patient study number:

Date of assessment:

Date of birth:

Sex:
1 = Male
2 = Female

Occupation

Side:
1 = Right
2 = Left

Symptoms

Pain:
4 = Never
3 = Rarely
2 = Occasionally
1 = Constantly

Instability:
4 = Never
3 = Rarely
2 = Occasionally
1 = Constantly

Locking:
4 = Never
3 = Rarely
2 = Occasionally
1 = Constantly

Swelling:
4 = Never
3 = Rarely
2 = Occasionally
1 = Constantly

Function

Stairs:
4 = Normal
3 = Difficult
2 = With Aid
1 = Unable

Sitting:
4 = Normal
3 = Difficult
2 = With Aid
1 = Unable

Sports:
4 = Normal
3 = Difficult
2 = With Aid
1 = Unable

Signs

Crepitus:
1 = Yes
0 = No

Apprehension:
1 = Yes
0 = No

Tenderness:
1 = Yes
0 = No

Q angle:
1 = Yes
0 = No

R.O.M.

Flexion:

Extension:

Strength Quadriceps:
M.R.C. 0-5

X-rays

Wiberg Type:
4 = Normal
3 = Better
2 = Same
1 = Worse

PT/PL ratio:
4 = Normal
3 = Better
2 = Same
1 = Worse

O.A-P.F.J.:
4 = Normal
3 = Better
2 = Same
1 = Worse

O.A.-Knee:
4 = Normal
3 = Better
2 = Same
1 = Worse

Patient Assessment :

4 = Normal
3 = Better
2 = Same
1 = Worse

Patella-femoral Trial

Treatment Assessment

Patient study number

Date of Assessment

Date of operation

Diagnosis

Operation

Lateral release

Lateral release & Tibial Tussicle Tunnel

Carbon pad arthroplast

Isolated new patello-femoral joint

Patella-femoral Trial

Final Assessment

Patient study number:

Date of post-operative assessment:

Symptoms

Pain:
- 4 = Never
- 3 = Rarely
- 2 = Occasionally
- 1 = Constantly

Instability:
- 4 = Never
- 3 = Rarely
- 2 = Occasionally
- 1 = Constantly

Locking:
- 4 = Never
- 3 = Rarely
- 2 = Occasionally
- 1 = Constantly

Swelling:
- 4 = Never
- 3 = Rarely
- 2 = Occasionally
- 1 = Constantly

Function

Stairs:
- 4 = Normal
- 3 = Difficult
- 2 = With Aid
- 1 = Unable

Sitting:
- 4 = Normal
- 3 = Difficult
- 2 = With Aid
- 1 = Unable

Sports:
- 4 = Normal
- 3 = Difficult
- 2 = With Aid
- 1 = Unable

Signs

Crepitus:
- 1 = Yes
- 0 = No

Apprehension:
1 = Yes
0 = No

Tenderness:
1 = Yes
0 = No

Q angle:
1 = Yes
0 = No

R.O.M.

Flexion:

Extension:

Strength Quadriceps:
M.R.C. 0-5

X-rays

Wiberg Type:
4 = Normal
3 = Better
2 = Same
1 = Worse

PT/PL ratio:
4 = Normal
3 = Better
2 = Same
1 = Worse

O.A-P.F.J.:
4 = Normal
3 = Better
2 = Same
1 = Worse

O.A.-Knee:
4 = Normal
3 = Better
2 = Same
1 = Worse

Patient Assessment:
4 = Normal
3 = Better
2 = Same
1 = Worse

Post-operative complications:

Name of Asssessor:

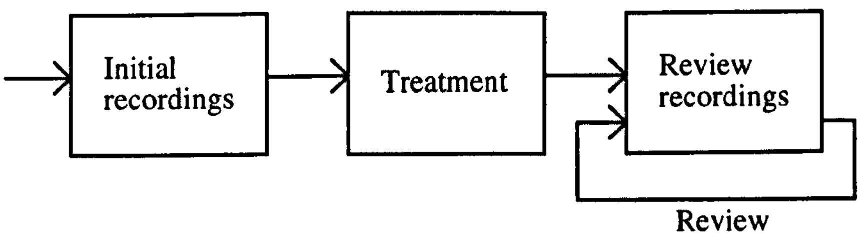
Appendix C (V)

Plant trial

This evaluation was designed to review the whole process of designing and building an IDDA end-system from the very beginning. In all the previous trials, the users were presented with a set of questionnaires from which their IDDA system was to be built. In this trial, the user (a biologist) was to construct an end-system from a series of questionnaires which he had also devised.

The trialist was a biologist, who had completed a year long M.Sc. computing conversion course, and his trial involved recording the growing patterns of wheat under varying conditions, for example, determining the outcome of using Absis Acid or Indol Acetic Acid, etc., in conjunction with differing lengths of light exposure. The effects were to be measured in terms of height, number of leaves, dry mass of a leaf, leaf colour and leaf condition (see pages C-27 through to C-32 for the questionnaires of the 3 stages). It was a simple, small experiment similar to those commonly run in biology laboratories before more expensive and elaborate trials are undertaken.

Three stages were required:



with reviews taking place every week for up to 7 months, which is the growing period of wheat.

Points to note

The production of the assessment questionnaires took an hour, which included the initiation of the trial, the specification of the questions at each stage, the entry of these questionnaires into a word processor, and, the production of the ASCII files. Following a similar process, the two sections of help text took 20 minutes (see page C-33).

The IDDA system was successfully constructed. The time taken to build the IDDA system was 30 minutes, which included the automatic construction of the end-system by the tools.

Overall the user found the tools very easy to use and felt quite confident after the first couple of screens. However, a number of difficulties arose during the construction because the help documentation had not been read prior to starting the actual trial. Consequently, not all the necessary planning had been undertaken. For example, the questionnaires were never written down on paper and therefore could not be referenced during the construction tasks. Instead, the user had to rely solely on his memory when attempting to recall details of previous or subsequent questions in the questionnaires. Nevertheless the IDDA end-system was built successfully and the difficulties only seem to have effected the time taken to complete the construction of the end-system as well as to have caused minor annoyance to the user.

The help system was referenced frequently throughout the construction phase. However, the operator recorded that he had 'forgotten' that the second monitor was there. This situation would only be a problem if the user is waiting for a response on the main screen when it had in fact appeared on the other display. As the second screen is only used for help text during the construction phase and the assumption is that a user who requests help needs extra information prior to progressing, it is highly likely that the user would look around for a response, thereby noticing the second screen. In this trial, the problem is likely to have occurred because the second monitor was placed lower than, and to the left of, the main monitor. Consequently, it was not in the normal range of vision of the user and could thus be easily overlooked. During previous trials, the screen had been placed adjacent to the main monitor on the right hand side and in the operator's normal field of vision. This position did not seem to cause any problems and therefore, in subsequent trials, this set-up will be used.

This trialist did not like the colours selected for the interface screens of the IDDA end-system. A facility to enable operators to choose their own colours could be implemented. However, there is a concern that naive computer users will not be able to select reasonable colour combinations for their end-system. Moreover, what if a number of users disagree over the colours to choose, would such a facility cause more confusion? The colours of the interfaces were chosen after the literature had deemed them restful and unobtrusive. They have already been used in a number of other medical systems and, to date, there have been no complaints from the users of those systems. Hence, the views of subsequent trialists will be noted with interest but for the time being the colours will stay unaltered and the facility for the users to select their own colours will stay unimplemented.

Another note made by this trialist was the wish to have tools to 'maintain' a constructed or partially constructed IDDA system. This ability would certainly be necessary in a final package. However, the facilities to:

- redefine the definition for the current assessment question,
- stop at the current assessment question and then re-start from this point later,
- stop and re-start from the beginning again,
- add more help text at a later date,

are sufficient for these tools, which are primarily investigating the feasibility of constructing an IDDA system.

From the times and comments recorded in this trial, it does seem that the planning, writing, and typing of the questionnaires and the help text are going to be the most demanding aspects of building an IDDA system. If these phases are carried out diligently, the actual operation of the tools is very easy and the construction of the IDDA end-system is both straightforward and logical. In some ways, forcing the builder to spend time over the planning stages is beneficial as it is likely to encourage deeper thought and consideration over the questionnaires and the subsequent assessment procedure as a whole. This should result in a better system being built and ultimately a 'better' experiment being carried out, i.e. 'better' in terms of being more justifiable and worthwhile.

Plant Trial

Initial Recordings Questionnaire

Plant Type:

Date of Initial Assessment:

Average ht (cm):

Average number of leaves:

Average Dry Mass (g):

Leaf Colour:

Leaf Appearance:
1 - Eaten
2 - Mottled
3 - Normal

Plant Trial

Initial Recordings (as typed by evaluator)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is question asked?
1	Plant Type	N	0	999999	0	-----	0	Always
2	Date initial rec.	D	-----	-----	-----	-----	<	Always
3	Average ht (cm)	N	2	999	0	-----	0	Always
4	Average no. leaves	N	0	999	0	-----	0	Always
5	Average dry mass	N	2	999	0	-----	0	Always
6	Leaf colour	A	-----	-----	-----	30	-----	Always
7	Leaf appearance	N	0	3	1	-----	3	Always

Plant Trial

Treatment Questionnaire

Plant Type:

Date of Treatment:

Treatment:

1 - Control
2 - ABA
3 - IAA
4 - DDT

Lighting:

1 - Normal 12 + 12
2 - Long Day 18 + 6
3 - Short Day 6 + 18

Plant Trial

Treatment (as typed by the evaluator)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Common Answer	When is question asked?
1	Plant Type	N	0	999999	0	0	Always
2	Date of treatment	D	-----	-----	-----	<	Always
3	Treatment	N	0	4	1	1	Always
4	Lighting	N	0	3	1	1	Always

Plant Trial

Final Recordings Questionnaire

Plant Type:

Date of Final Assessment:

Average ht (cm):

Average number of leaves:

Average Dry Mass (g):

Leaf Colour:

Leaf Appearance:
1 - Eaten
2 - Mottled
3 - Normal

Plant Trial

Final Recordings (as typed by the evaluator)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is question asked?
1	Plant Type	N	0	999999	0	-----	0	Always
2	Date final rec.	D	-----	-----	-----	-----	<	Always
3	Average ht (cm)	N	2	999	0	-----	0	Always
4	Average no. leaves	N	0	999	0	-----	0	Always
5	Average dry mass	N	2	999	0	-----	0	Always
6	Leaf colour	A	-----	-----	-----	30	-----	Always
7	Leaf appearance	N	0	3	1	-----	3	Always

Plant Trial

Help sections specified by the evaluator

Leaf Colour.

Enter the predominant colour of a leaf. If there are two colours that are equally dominant, enter both with a '/' dividing them. If there are more than two colours of equal coverage, enter 'mixed'.

||END||

Leaf Appearance.

'Eaten' is when a leaf has been eaten by a animal or insect but not when it has decayed or been damaged by other forces.

||END||

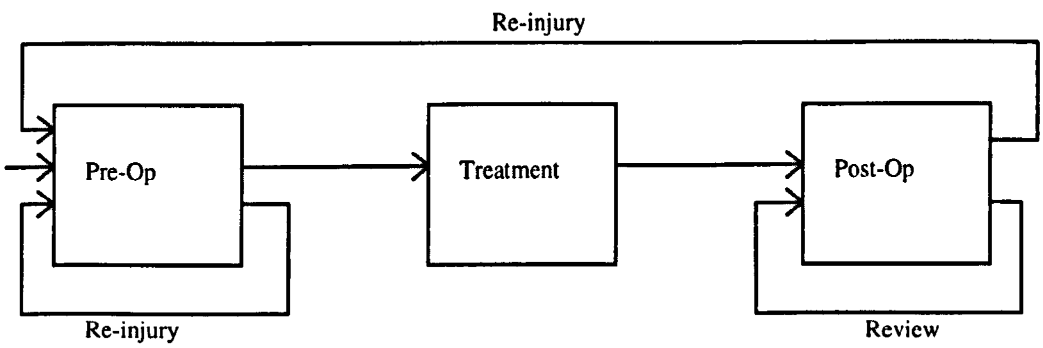
Appendix C (VI)

Patella-femoral joint trial (II)

This investigation was to determine whether secretarial staff, who were unfamiliar with the specialist field and who were naive computer users, could successfully construct a specified IDDA system using the tools. If successful, could this approach be used rather than the specialist spending time constructing his system? The specialist, however, would still be needed to produce the required assessment questionnaires. Once again, the interface and functionality of the tools were also being reviewed as well as the correctness of the code produced by the tools.

Two secretarial staff were selected for this study. One had spent 6 months becoming familiar with a word-processor, as a part-time secretary. Previous to that she had no experience of computing and had not used computers before. The other had carried out a data entry role in a previous job but other than word-processing in her current position as a part-time secretary, she had very limited experience of computers or computing.

The patella-femoral assessment procedure was used for the trial. As described in a previous trial (Appendix C (IV)), this method is currently carried out on patients who consistently complain about 'pain behind the kneecap'. It consists of questionnaires undertaken in three stages - Pre-operative, Treatment, and Post-operative.



These assessment questionnaires were derived from the original assessment form (see Appendix C (IV)) by Mr Harding an orthopaedic consultant at Leicester General Hospital.

Both the secretaries typed the questions into files, for each of the three stages, from hand-written information. No problems occurred during the completion of this task. As explained above, both have gained experience of using a word-processor and were therefore able to utilise this previously gained knowledge.

They did however differ in the way they set out the information - one placed a blank line after each question before listing the options available for that question; the other varied her approach, though more often she did not use a blank line. Both used tabs to indent their information rather than space-characters. The presented hand-written information did not have any blank lines between the questions and the option lists, although they did have the option lists indented.

In this trial there was no help specified to accompany the questionnaires, therefore the secretaries did not need to construct any help files.

Details for each of the questions in the three stages were given to the secretaries in hand-written form. One version had these details laid out in a sequence appropriate to the questions the generator would ask; the other version contained all the required information in an ordered form but not necessarily in the exact sequence for the generator's questions (see pages C-38 through to C-43). Doctors are not renowned for the structure of their reporting, although generally all the information is presented somewhere. The second version was therefore used to accommodate this fact, i.e. that the information given to the secretaries by the doctors might not be laid out in an ideal way.

Observations noted during the trial:

Secretary One (only limited word-processing experience; information not set out in the ideal format)

At the start she was rather apprehensive as she did not know the specialist area for which the IDDA system was being built. In addition, it was the first time she had encountered a computer system asking her questions and certainly the first time a computer had asked her to confirm her entries.

The above problem was exacerbated by the fact that she perceived that she should know the answers without having to refer to the doctor's notes. Gradually she realised this was not the case, although it was not until the Stage Three questionnaire that she referenced the notes as a natural response to any uncertainties.

Having said this, it was noticeable that she seemed to be happier and to begin to speed up her response rate after question 7 in Stage One. However, switching to specifying a conditional question did cause a problem initially. A conditional question is one that requires a condition to be defined for when the question is asked, i.e. the asking of the question is reliant on answers obtained from previous questions. There were only three questions requiring conditions in this trial. By the time she was defining the third she, again, seemed happy translating the information specified to the required condition. Having to refer to the doctor's notes at this point (the end of Stage Two) might be the reason why she began to use the doctor's notes more frequently as a reference during Stage Three.

A disconcerting note was the fact that she did begin to respond to the questions automatically as her confidence grew, with no checking of her responses with the notes. Certainly some of the information can, at times, be gleaned from the presented question, e.g. the maximum or minimum numeric entry allowed. However, other information such as the common answer or when the question is to be asked is not evident from the information displayed. She seemed to be entering what she thought the answer should be and not what the doctor had indicated it as being. Most of the time she was, in fact, correct, and some of the other times the computer asked her to confirm her response which did prompt her to check. If she actually worked within the specialist field, she might well have been correct all of the time, but this is, of course, not certain.

At no time during the generation of the IDDA did she request any help from the system.
The time taken was 35 minutes. This time included the production and compilation of all the programs for the defined IDDA end-system, e.g. it took 35 minutes to construct a fully working system once the questionnaires had been defined on paper and been typed in via a word-processor.

Secretary Two (some data entry experience; layout more appropriate to the sequence of questions asked)

This trialist was certainly happier with the machine driving the dialogue, e.g. the machine asking her questions. Moreover, she used the doctor's notes extensively and pedantically.

As before, specifying the conditional questions initially caused some problems. This again could be due to the lack of knowledge of the specialist field. However, by the third condition she appeared happier with the process of defining a condition, although not as comfortable as the first secretary. These observations seem to indicate that more detailed explanation of conditions, or more practice, is required prior to using the generator (if, of course, there are any conditional questions to be defined).

At no time during the generation of the IDDA did she request any system help. The system that was produced worked precisely as the doctor had specified. It took 45 minutes to construct, including the automatic generation, after the questionnaires had been typed in using a word-processor.

Points to note from the trial:

The time taken to construct the end-system was not long, even when the operators had no knowledge of the specialist field. It seems likely therefore that it would require less time to build the system if the users do know the field as they would not need to reference the written notes as often, but it is unlikely that the task could be completed substantially faster. Moreover, the preparation and layout of the stage questionnaires would take longer to accomplish if specialists have to clearly explain their requirements to the user of the tools. Therefore, since the construction time is so short, the writing of the notes in a neat and ordered fashion for someone else is highly likely to take longer than merely noting the information down and running the generator yourself.

In addition, the generator requires little typing, which is the primary skill that the secretaries can offer. Most of the entries are merely selections from presented lists. Consequently, speed of entry and touch typing are not of major importance for the successful operation of the generator tools.

Therefore, there would be little, if any, time saved through having the secretarial staff construct the end-system rather than, as in this case, the doctors carrying out the task themselves. Furthermore, there is the facility to 'Quit and Save' during the operation of the tools, which permits a user to leave and return later to continue the construction task. This facility is especially useful if the assessment questionnaires are lengthy or if the user is called away unexpectedly.

With a specialist or a junior member of the team as the operator, the users will be aware of the specialist area, the devised assessment procedures, the terminology and the responses used, and, in fact, the whole process the system is attempting to mimic or portray. This knowledge certainly assists in the problematic area of defining conditional questions, especially if users themselves have devised, or have assisted in, the production of the questionnaires.

Finally, if specialists, or members of the team, have operated the generator to build the end-system, they will be much more confident that the produced system has actually been defined correctly and will operate as they require. This ability to have confidence in a computer system is crucial to obtaining user acceptance of an end-system.

Consequently, after considering just this last point, the recommendation will be that the specialist or a member of the team should operate the generator tools to construct the required system, thereby ensuring at least some confidence in the final system at the possible expense of spending a small amount of extra time constructing it.

Patella-femoral Trial

Pre-Operative Assessment (Secretary 1)

Question Number	Question	Type	Range	Dec. Places	Common Answer	When is the question asked?
1	Patient Study	N	0 - 999999	0	0	Always
2	Date of Ass.	D	-------------	-----	< (before)	Always
3	Date of Birth	D	-------------	-----	< (before)	Always
4	Sex of Patient	N	1 - 2	0	1	Always
5	Occupation	A	30 spaces	-----	(nothing)	Always
6	Side	N	1 - 2	0	1	Always
7	**Symptoms**	C	-------------	------	--------	**<Next question>**
7	Pain	N	1 - 4	0	4	Always
8	Instability	N	1 - 4	0	4	Always
9	Locking	N	1 - 4	0	4	Always
10	Swelling	N	1 - 4	0	4	Always
11	**Function**	C	-------------	------	--------	**<Next question>**
11	Stairs	N	1 - 4	0	4	Always
12	Sitting	N	1 - 4	0	4	Always
13	Sports	N	1 - 4	0	4	Always
14	**Signs**	C	-------------	------	--------	**<Next question>**
14	Crepitus	N	0 - 1	0	0	Always
15	Apprehension	N	0 - 1	0	0	Always
16	Tenderness	N	0 - 1	0	0	Always
17	Q-Angle	N	0 - 1	0	0	Always
18	**R.O.M.**	C	-------------	------	--------	**<Next question>**
18	Flexion	N	0 - 180	2	0	Always
19	Extension	N	0 - 180	2	0	Always
20	Strength Quad.	N	0 - 5	0	0	Always
21	**X-Rays**	C	-------------	------	--------	**<Next question>**
21	Wiberg Type	N	1 - 4	0	4	Always
22	PT/PL ratio	N	1 - 4	0	4	Always
23	O.A. - P.F.J.	N	1 - 4	0	4	Always
24	O.A. - Knee	N	1 - 4	0	4	Always
25	Patient Ass.	N	1 - 4	0	4	Always

Patella-femoral Trial

Treatment Assessment (Secretary 1)

Question Number	Question	Type	Range	Common Answer	When is the question asked?
1	Patient Study	N	0 - 999999	0	Always
2	Date of Ass.	D	------	<	Always
3	Date of Op.	D	------	<	Always
4	Diagnosis	A	30 (spaces)	(nothing)	Always
5	**Operation**	**C**	**------**	**------**	**<Next question>**
5	Lateral release	L	Y/N	N	Always
6	Lat. rel. and tib.	L	Y/N	N	5 = 'N'
7	Carbon Fibre	L	Y/N	N	5 = 'N' and 6 = 'N'
8	Isolated new Pat.	L	Y/N	N	5 = 'N' and 6 = 'N' and 7 = 'N'

Patella-femoral Trial

Post-Operative Assessment (Secretary 1)

Question Number	Question	Type	Range	Dec. Places	Common Answer	When is the question asked?
1	Patient Study	N	0 - 999999	0	0	Always
2	Date of Ass.	D	-------------	-----	< (before)	Always
3	**Symptoms**	**C**	-------------	------	--------	**<Next question>**
3	Pain	N	1 - 4	0	4	Always
4	Instability	N	1 - 4	0	4	Always
5	Locking	N	1 - 4	0	4	Always
6	Swelling	N	1 - 4	0	4	Always
7	**Function**	**C**	-------------	------	--------	**<Next question>**
7	Stairs	N	1 - 4	0	4	Always
8	Sitting	N	1 - 4	0	4	Always
9	Sports	N	1 - 4	0	4	Always
10	**Signs**	**C**	-------------	------	--------	**<Next question>**
10	Crepitus	N	0 - 1	0	0	Always
11	Apprehension	N	0 - 1	0	0	Always
12	Tenderness	N	0 - 1	0	0	Always
13	Q-Angle	N	0 - 1	0	0	Always
14	**R.O.M.**	**C**	-------------	------	--------	**<Next question>**
14	Flexion	N	0 - 180	2	0	Always
15	Extension	N	0 - 180	2	0	Always
16	Strength Quad.	N	0 - 5	0	0	Always
17	**X-Rays**	**C**	-------------	------	--------	**<Next question>**
17	Wiberg Type	N	1 - 4	0	4	Always
18	PT/PL ratio	N	1 - 4	0	4	Always
19	O.A. - P.F.J.	N	1 - 4	0	4	Always
20	O.A. - Knee	N	1 - 4	0	4	Always
21	Patient Ass.	N	1 - 4	0	4	Always
22	Patient Comp.	A	30 (spaces)	------	<blank>	Always
23	Name Assessor	A	30 (spaces)	------	<blank>	Always

Patella-femoral Trial

Pre-Operative Assessment (Secretary 2)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	999999	0	-------	0	Always
2	Date of Ass.	D	-----	------	------	-------	< (before)	Always
3	Date of Birth	D	-----	------	------	-------	< (before)	Always
4	Sex of Patient	N	0	2	1	-------	1	Always
5	Occupation	A	-----	------	------	30 spaces	(nothing)	Always
6	Side	N	0	2	1	-------	1	Always
7	**Symptoms**	**C**	------	------	------	-------	--------	**<Next question>**
7	Pain	N	0	4	1	-------	4	Always
8	Instability	N	0	4	1	-------	4	Always
9	Locking	N	0	4	1	-------	4	Always
10	Swelling	N	0	4	1	-------	4	Always
11	**Function**	**C**	------	------	------	-------	--------	**<Next question>**
11	Stairs	N	0	4	1	-------	4	Always
12	Sitting	N	0	4	1	-------	4	Always
13	Sports	N	0	4	1	-------	4	Always
14	**Signs**	**C**	------	------	------	-------	--------	**<Next question>**
14	Crepitus	N	0	1	0	-------	0	Always
15	Apprehension	N	0	1	0	-------	0	Always
16	Tenderness	N	0	1	0	-------	0	Always
17	Q-Angle	N	0	1	0	-------	0	Always
18	**R.O.M.**	**C**	------	------	------	-------	--------	**<Next question>**
18	Flexion	N	2	180	0	-------	0	Always
19	Extension	N	2	180	0	-------	0	Always
20	Strength Quad.	N	0	5	0	-------	0	Always
21	**X-Rays**	**C**	------	------	------	-------	--------	**<Next question>**
21	Wiberg Type	N	0	4	1	-------	4	Always
22	PT/PL ratio	N	0	4	1	-------	4	Always
23	O.A. - P.F.J.	N	0	4	1	-------	4	Always
24	O.A. - Knee	N	0	4	1	-------	4	Always
25	Patient Ass.	N	0	4	1	-------	4	Always

Patella-femoral Trial

Treatment Assessment (Secretary 2)

Question Number	Question	Type	Max. Answer	Min. Answer	Log. Type	Length	Common Answer	When is the question asked?
1	Patient Study	N	999999	0	------	------	0	Always
2	Date of Ass.	D	------	------	------	------	< (before)	Always
3	Date of Op.	D	------	------	------	------	< (before)	Always
4	Diagnosis	A	------	------	------	30 (spaces)	(nothing)	Always
5	**Operation**	**C**	------	------	------	------	------	**<Next question>**
5	Lateral release	L	------	------	1	------	N	Always
6	Lat. rel. and tib.	L	------	------	1	------	N	5 = 'N'
7	Carbon Fibre	L	------	------	1	------	N	5 = 'N' and 6 = 'N'
8	Isolated new Pat.	L	------	------	1	------	N	5 = 'N' and 6 = 'N' and 7 = 'N'

Patella-femoral Trial

Post-Operative Assessment (Secretary 2)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	999999	0	-------	0	Always
2	Date of Ass.	D	-----	------	------	-------	< (before)	Always
3	**Symptoms**	C	------	------	------	-------	--------	**<Next question>**
3	Pain	N	0	4	1	-------	4	Always
4	Instability	N	0	4	1	-------	4	Always
5	Locking	N	0	4	1	-------	4	Always
6	Swelling	N	0	4	1	-------	4	Always
7	**Function**	C	------	------	------	-------	--------	**<Next question>**
7	Stairs	N	0	4	1	-------	4	Always
8	Sitting	N	0	4	1	-------	4	Always
9	Sports	N	0	4	1	-------	4	Always
10	**Signs**	C	------	------	------	-------	--------	**<Next question>**
10	Crepitus	N	0	1	0	-------	0	Always
11	Apprehension	N	0	1	0	-------	0	Always
12	Tenderness	N	0	1	0	-------	0	Always
13	Q-Angle	N	0	1	0	-------	0	Always
14	**R.O.M.**	C	------	------	------	-------	--------	**<Next question>**
14	Flexion	N	2	180	0	-------	0	Always
15	Extension	N	2	180	0	-------	0	Always
16	Strength Quad.	N	0	5	0	-------	0	Always
17	**X-Rays**	C	------	------	------	-------	--------	**<Next question>**
17	Wiberg Type	N	0	4	1	-------	4	Always
18	PT/PL ratio	N	0	4	1	-------	4	Always
19	O.A. - P.F.J.	N	0	4	1	-------	4	Always
20	O.A. - Knee	N	0	4	1	-------	4	Always
21	Patient Ass.	N	0	4	1	-------	4	Always
22	Patient Comp.	A	------	------	------	30 (spaces)	(nothing)	Always
23	Name Assessor	A	------	------	------	30 (spaces)	(nothing)	Always

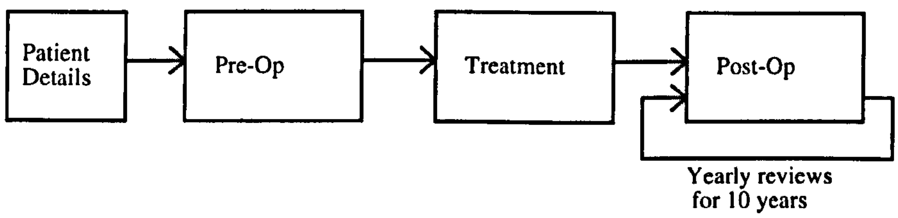
Appendix C (VII)

Comparing two Knee Replacement systems and specifying the benefits of such systems

The purpose for this trial was two-fold: firstly, to compare a system that had been built by the IDDA tools with a system that had been built by hand by another developer; and secondly, to obtain the views of a medical consultant regarding the benefits that he perceives such systems could bring to specialist medical fields.

The application area selected for the comparison concerned knee replacements (see pages C-50 through to C-59 for the questionnaires and the answer details for the questions in the questionnaires). A computer-based system had been developed for the Glenfield General Hospital in Leicester by D.K. Bhatt. It was written in dBASE III+ and used the statistical package SPSS/PC+ for data analysis.

Consisting of three databases, it collects and stores patient details for yearly check-ups over the 10 year period after an operation. Patients can be uniquely selected to enable their details to be viewed on the screen or printed out on the local printer. The databases were set-up in the following manner:



For statistical analysis, the developer determined that the medical consultant should convert the databases to ASCII text before loading these into SPSS/PC+. The consultant could then specify the test and the data items under investigation. Though they have the full power of the statistical package, the consultants only used frequencies, means, t-tests, variance analysis and bar charts. A sample of the output they currently receive can be seen on pages C-60 to C-67 of this Appendix.

i) The IDDA system

The system constructed by the IDDA tools modelled the original questionnaires devised by the orthopaedic consultants. These were broken down into three stages consisting of 36 questions in Stage One, 5 questions in Stage Two, and 31 questions in Stage Three. The patient's name and address were held separately.

With the IDDA knee replacement system, the consultants can carry out all of the analyse they require without having to convert the data to ASCII text, quit the system, initiate a statistical package, and in load the data. A further benefit is that they do not have to remember the field names of the data items they wish to use in their analysis. They only need to select the appropriate question numbers from the required questionnaire, after specifying the test they wish to carry out, and the system will then produce the query. Consequently this method should save consultants both time and effort.

Points to note:

a) The previous knee replacement system:

Mr Oni, the orthopaedic consultant involved in the original project, explained the short-comings of the initial system built by D.K. Bhatt as:

a) the system had no in-built methods for analysing data - the doctors could not see the sense of collecting the data in a database and then extracting it and loading it into a different package for analysis,
b) there was no means by which to select the data to be 'dumped', e.g. all the data in a database was written to an ASCII text file,
c) the statistical package SPSS was too complex for the doctors to operate,
d) the printing facility only permitted the selection of all patients - you could not specify one patient and just receive the one page printout,
e) the screens were densely packed with information.

The result of experiencing these problems has been a reluctance on the part of the doctors to use the system. Instead, the doctors are selecting other ad-hoc methods to gather their data and to undertake data analysis in an attempt to overcome the difficulties of the computer system, i.e. the system is now rarely used.

Some of these difficulties could have arisen from the lack of consultation with Mr Oni during the development process. Mr Oni recalls that:

a) there was no consultation during the design or development,
b) the original proposal had been to build a system that the orthopaedic secretaries could use, to enter and store the patient details. However, there was no discussion between the developer and secretaries to try to understand the daily tasks that secretaries undertake, nor the environment in which they work nor their knowledge of computers or the specialist medical field,
c) the anticipated result of having the system was that it 'would make life easier for the doctor' - it was felt that the system had not achieved this goal,
d) the system took a full year to develop.

Overall, Mr Oni was rather disappointed with this original project, especially with the lack of consultation and his inability to contribute or participate. He had had a number of ideas that he would have liked to explore, and would like to have been able to comment on the system as the development progressed, e.g. he felt left out of the whole process even though he had originally initiated the project.

Since the doctors found SPSS too complex, they currently use Minitab for data analysis. They find this much easier to operate. They define the columns and headings, then add the data into the relevant column by hand before selecting the relevant columns to be compared. However, this method means that for each of the different treatment strategies and for each review assessment made, i.e. 6, 12, 18, 24 weeks, there needs to be a separate column. This results in a very large table being produced relatively quickly and requires a substantial amount of time for entering the different data items.

For a simple determination of the mean, standard deviation and the number of entries included in the columns, the doctors currently use a separate spreadsheet package. Again, they define the column, the headings and enter the data by hand. They opted to use a spreadsheet as they found it could be easily integrated into a report or a scientific paper.

These methods are used for every analysis a doctor wishes to undertake. They are error-prone, as there is no check on the data values as they are entered, time-consuming and long-winded.Doctors found that they rarely had enough time to carry out all the analyses that they would have liked.

This resulted in the doctors believing that research requiring such investigations and data analysis was too tedious to undertake. If anyone was interested in these types of study, the common advice was to 'pack up your results and post them to a statistician and just be prepared to pay'.

b) The IDDA system built by the tools

The primary complaint directed at the original system (built by D.K. Bhatt) seemed to stem from the lack of consultation that had occurred during the development process. This resulted in the building of an inappropriate system for the intended users and tasks. In contrast, since the doctors themselves construct the product, the IDDA system built by the tools has the important characteristic of being moulded by the working practices and the environment. Mr Oni viewed this as highly beneficial as it enabled the doctors to control the development process, in terms of time, content and action.

Another major omission was the lack of appreciation of the importance of being able to analyse the collected data easily, quickly and within the same system. As most medical studies require statistical analysis of the results, this feature was deemed by Mr Oni as being an essential component of any medical computer-system. Therefore, when the IDDA system built by the tools was demonstrated to Mr Oni, he was most impressed with the in-built data analysis capabilities. This, it seems, was exactly what he had anticipated with the original project. The ability for data to be checked as it is entered and the facility to review individual patient case-histories were also deemed especially useful. The simplicity of operation, he claimed, enabled him to feel happy and confident in allowing the secretaries to undertake the data entry, thus releasing him to carry out further investigations.

Mr Oni perceived that the current ability for consultants to build their own system, thereby tailoring it to the study, was a major step forward. This was mainly because this approach enabled the consultants to be in charge of developing their systems and permitted them to specify when a new system was to be developed. This control also stretched to determining the time it took to construct the system, the content and the operation of the system.

Moreover, the number of medical studies constantly being initiated makes this facility even more important. The current desire by the medical profession to attempt to share data, thereby increasing the sample size quickly, has resulted in the realisation that there must be a drive towards standardising procedures and the terminology used within specialist fields. The standardisation may not include all the practices but the data that is shared must have been collected in the same manner to be of any value. Therefore there needs to be detailed discussions to debate which techniques should be used as 'standard' and which are not required. This indicates a need to carry out comparative analyses on the various alternatives. Such investigations, it was felt, could be carried out using IDDA systems.

Two additional facilities that Mr Oni thought might be useful were:

a) the ability to turn off the requirement for a second monitor to be attached.

   This would enable the end-system to run on a portable computer, thus it could be used in field studies. The role of the second screen is to display help and examples during the end-system's creation and during statistical analysis. Any other help produced by the doctors to be linked to the questions in the assessment questionnaires would also be displayed on the second monitor. Since it is highly likely that doctors undertaking remote field trials are experienced in the methods used to gather any data required, it is unlikely that they will need help to explain such techniques. In addition, the end-system will certainly have been constructed prior to setting out on the field trial (therefore there will be no requirement for the construction help). Consequently, the only help that might be of use in this scenario would be the analysis help.

   In general, however, the aim of the field trial is to gather the data rather than to analysis it 'on the hoof', thus the requirement to carry out statistical analysis whilst away from the main site is unlikely. Moreover, once back at the main centre, the necessity for not having a second monitor diminishes. Therefore it could be there that any analysis on the collected data is conducted and where any new systems for further field trials are constructed.

   Hence there could well be a situation where attaching a second monitor is impractical. Thus, the facility to be able to select whether or not an end-system is to have access to a second screen is a reasonable addition. To achieve this, the tools will need to be changed but not extensively. Consequently, there is no reason why this facility should not be added when the tools are next updated.

b) the ability to have a 'glance' chart.

   This is to allow certain selected questions to be part of a one page sheet that is divided into columns to represent various different reviews at different times, i.e.

Review patient progress

Patient study number: 126

	1st review	2nd review	3rd review
Date of review	12/09/90	20/10/01	25/10/92
Freq. Pain (1-5)	3	3	2
Sev. Pain (1-5)	4	3	3
Night Pain (1-5)	2	2	2

Press any key to continue..........

Currently within the tools there is the ability to define a table during the construction of an end-system. However this table would be included in the data entry phase of the end-system, i.e. included as part of an assessment questionnaire. It should be noted that the 'table' facility as well as the 'list' facility have, as yet, never been used during any of the evaluations. Consequently, a change to the tools could include removing these two facilities and to add the facility of selecting questions to be part of a 'glance' chart. This new facility could make extensive use of the code currently available to construct a table and therefore the inclusion of this new facility should not cause too many problems.

As yet, Mr Oni has been the only evaluator to request this ability. Therefore the only other consideration would be whether it would be better to have a 'glance' chart facility as an add-on extra, which would be included if the user requests, rather than as a standard part of the tools. There will have to be further discussions with medical consultants, as well as exposing the tools to other situations and environments, to attempt to determine which of these two methods is the better.

Therefore, both of these suggested additions will be considered when reviewing future improvements to the tools and to the general IDDA end-system.

ii) Benefits of such computer-based systems

Mr Oni perceived that clinical care would be improved by using computer-based systems similar to the IDDA end-system. He listed a number of reasons for this belief, which included:

a) the ability that such systems gave to the doctors to easily manipulate their data,
b) the encouragement that these systems give to the doctors to undertake further research. This is due, in the main, to such systems making investigations more detailed and in-depth, as well as quicker and easier to carry out, whilst also making them less tedious and less error-prone,
c) the ability to undertake audits of the practices and strategies used by different research sites, which could range from being world-wide to simply within a single department,
d) being able to review the outcomes from the results from c), concerning which practices and tests should be used and which are of little benefit. This would lead to an agreed standardised strategy for dealing with patients, as well as reducing: the time currently taken in the assessment and treatment procedures; the various monetary costs; and, the actual discomfort experienced by patients,
e) the realisation of the ultimate aim of producing more formalised specialist fields with clearer insights and a deeper understanding of the whole medical domain and the various ailments.

With regards to the IDDA tools, Mr Oni believed that they could assist in achieving all of the above. For example:

a) the ability the tools give the doctor to build his/her own system and thereby enabling the development of an appropriate system for the working environment and in the time scale dictated by the doctors themselves,

b) the speed with which these end-systems could be constructed using the tools, thus enabling new studies to be initiated with the minimum of delay,

c) the in-built analysis and review capabilities of the end-system produced by the tools, enabling the doctors to complete their investigations within one system.

In addition, Mr Oni perceived that specialist computer-based end-systems produced by the IDDA tools would not interfere with the patient-physician relationship nor would they delay work unduly, whilst they would fit easily into the working practices of the orthopaedic unit.

Current computer systems that are available could, in his opinion, be improved by enabling the 'computerisation of the design and production of assessments charts as well as the automatic production of reports or statistics, etc., without requiring a "middle-man" or a degree in computing or statistics.' As Mr Oni had this preconceived role for computers in the medical environment, he was most impressed with the capabilities of the demonstrated tools and the end-system they produced.

He perceives that the purpose of computer systems in specialist medical fields would be for auditing and research. He is personally involved in both of these areas and is therefore well aware of the benefits that appropriate computerisation could bring. He has asked if he could use the demonstrated tools if his latest research grant for a study of 300 patients over two years is accepted.

Moreover, he proposes to contact Johnson and Johnson who are currently looking for a centre to base the analysis of an international study on knee replacement operations. Johnson and Johnson wish to computerise the data collection at a number of centres world-wide and then amalgamate these details in one central computer system on which statistical analyses will be undertaken. The reason for this study is to carry out an audit of all the operations and treatments carried out at the different centres and to try to arrive at some standardisation for the recording of the information and to try to rationalise the number of tests undertaken. Moreover, it is hoped that the study will indicate which are the better treatment strategies to adopt for patients who require knee replacement surgery. Mr Oni believes that the ability of the tools to quickly construct appropriate end-systems for each local centre as well as the integrated analysis capabilities, results in the tools being ideally suited for the planned study.

Consequently, from these discussions with Mr Oni and his responses to the questionnaires completed after the evaluation, he was impressed with the demonstrated IDDA tools and the end-system they produce. So much so that he has requested to use the IDDA tools in a future study. In addition, he believes the concept of doctors building their own end-system using the tools to be highly beneficial and a major step forward in the production of computer-based systems for the medical field.

Knee Replacement System

Initial Assessment

Study number of patient

Date of Assessment

Hospital Number

Date of Birth

Sex of patient

1 = Male
2 = Female

Side

1 = Left
2 = Right

Diagnosis

1 = oa
2 = ra
3 = osteon
4 = sa
5 = other

Other Diagnosis

Weight (Kg)

Frequency of Pain

1 = Pain all the time
2 = Slight pain at rest but goes when moving
3 = Most of time no pain
4 = Aches but gets better by itself
5 = Hardly/never hurts

Severity of Pain

1 = Very severe
2 = Really bad
3 = Quite bad
4 = Not very bad
5 = No pain

Pain at Night

1 = Every night awake for long periods
2 = Most nights awake for long periods
3 = Quite often
4 = Only occasionally
5 = Never

Walking Ability (patient walks 25 yards using normal aid)

1 = can not walk 25 yards
2 = More than 45 seconds
3 = In 33 - 45 seconds
4 = In 18 - 32 seconds
5 = Less than 18 seconds

Walking Aid

1 = Not applicable - can not walk 25 yards
2 = Frame
3 = Support both sides
4 = Support one side
5 = None

Sitting Down

1 = Needs human help
2 = Does not take fair share of strain
3 = With considerable help from arms
4 = With little help from arms
5 = Takes fair share of strain without help from arms

Rising Up

1 = Needs human help
2 = Does not take fair share of strain
3 = With considerable help from arms
4 = With little help from arms
5 = Takes fair share of strain without help from arms

Ability to Stand - affected knee bears whole body weight without support from hands, except for balance (best of 3 attempts)

1 = Less than 3 second or not at all
2 = 3 - 5 seconds
3 = 6 - 8 seconds
4 = 8 - 11 seconds
5 = 12 or more seconds

Going Upstairs - affected leg first and without help from arms

1 = can not step up 6 inches
2 = 6 inches high
3 = 8 inches high
4 = 10 inches high
5 = 12 inches high

Going Downstairs - affected leg second and without help from arms

1 = can not step up 6 inches
2 = 6 inches high
3 = 8 inches high
4 = 10 inches high
5 = 12 inches high

Coronal Tibiofemoral Angle for Valgus

Coronal Tibiofemoral Angle for Varus

Injured Knee

Range of movement

Flexion

Flexion Contracture

Extension Lag

Abduction hip

Foot RVF

Other Knee

Range of movement

Flexion

Flexion Contracture

Extension Lag

Abduction hip

Foot RVF

State of patella
1 = Normal function
2 = Palella-femoral pain
3 = Excised

Knee Replacement System

Pre-Operative Assessment (examples of the entries made)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	9999	0	-------	0	Always
2	Date of Ass.	D	-----	------	------	-------	< (before)	Always
3	Hospital Num.	A	------	------	------	7	(nothing)	Always
4	Date of Birth	D	-----	------	------	-------	< (before)	Always
5	Sex of Patient	N	0	2	1	-------	1	Always
6	Side	N	0	2	1	-------	1	Always
7	Diagnosis	N	0	5	1	-------	1	Always
8	Other Diag.	A	------	------	------	30	(nothing)	7 = 5
9	Weight	N	2	210	7	-------	70	Always
10	Freq. Pain	N	0	5	1	-------	5	Always
11	Sev. Pain	N	0	5	1	-------	5	Always
12	Pain at night	N	0	5	1	-------	5	Always
13	Walking ability	N	0	5	1	-------	5	Always
14	Walking Aid	N	0	5	1	-------	5	Always
15	Sitting Down	N	0	5	1	-------	5	Always
16	Rising up	N	0	5	1	-------	5	Always
17	Ability to stand	N	0	5	1	-------	5	Always
18	Going upstairs	N	0	5	1	-------	5	Always
19	Going downstairs	N	0	5	1	-------	5	Always
20	Cor. Tib. An. Val	N	2	60	0	-------	0	Always
21	Cor. Tib. An Var	N	2	60	0	-------	0	Always
22	**Injured Knee**	**C**	------	------	------	-------	------	**<Next question>**
22	Range movement	N	2	140	0	-------	0	Always
23	Flexion	N	2	140	0	-------	0	Always
24	Flexion Con.	N	2	90	0	-------	0	Always
25	Extension lag	N	2	180	0	-------	0	Always
26	Abduction Hip	N	2	60	0	-------	0	Always
27	Foot RVF	N	2	99	0	-------	0	Always
28	**Other Knee**	**C**	------	------	------	-------	------	**<Next question>**
28	Rang. movement	N	2	140	0	-------	0	Always
29	Flexion	N	2	140	0	-------	0	Always
30	Flexion Con.	N	2	90	0	-------	0	Always
31	Extension lag	N	2	180	0	-------	0	Always
32	Abduction Hip	N	2	60	0	-------	0	Always
33	Foot RVF	N	2	99	0	-------	0	Always
34	State Patella	N	0	3	1	-------	1	Always

Knee Replacement System

Treatment Assessment

Study Number of patient

Date of Assessment

Date of operation

Type of Operation

1 = Cemented
2 = Uncemented
3 = Cem. Femur/ Uncem. Tibia
4 = Uncem. Femur/ Cem. Tibia
5 = Other

Other Operation

Knee Replacement System

Treatment Assessment (examples of the entries made)

Question Number	Question	Type	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	9999	0	------	0	Always
2	Date of Ass.	D	------	------	------	< (before)	Always
3	Date of Op.	D	------	------	------	< (before)	Always
4	Type Op.	N	5	1	------	1	Always
5	Other Op.	A	------	------	30 (spaces)	(nothing)	4 = 5

Knee Replacement System

Final Assessment

Study number of patient

Date of Assessment

Weight (Kg)

Frequency of Pain

1 = Pain all the time
2 = Slight pain at rest but goes when moving
3 = Most of time no pain
4 = Aches but gets better by itself
5 = Hardly/never hurts

Severity of Pain

1 = Very severe
2 = Really bad
3 = Quite bad
4 = Not very bad
5 = No pain

Pain at Night

1 = Every night awake for long periods
2 = Most nights awake for long periods
3 = Quite often
4 = Only occasionally
5 = Never

Walking Ability (patient walks 25 yards using normal aid)

1 = can not walk 25 yards
2 = More than 45 seconds
3 = In 33 - 45 seconds
4 = In 18 - 32 seconds
5 = Less than 18 seconds

Walking Aid

1 = Not applicable - can not walk 25 yards
2 = Frame
3 = Support both sides
4 = Support one side
5 = None

Sitting Down

1 = Needs human help
2 = Does not take fair share of strain
3 = With considerable help from arms
4 = With little help from arms
5 = Takes fair share of strain without help from arms

Rising Up

1 = Needs human help
2 = Does not take fair share of strain
3 = With considerable help from arms
4 = With little help from arms
5 = Takes fair share of strain without help from arms

Ability to Stand - affected knee bears whole body weight without support from hands, except for balance (best of 3 attempts)

1 = Less than 3 second or not at all
2 = 3 - 5 seconds
3 = 6 - 8 seconds
4 = 8 - 11 seconds
5 = 12 or more seconds

Going Upstairs - affected leg first and without help from arms

1 = can not step up 6 inches
2 = 6 inches high
3 = 8 inches high
4 = 10 inches high
5 = 12 inches high

Going Downstairs - affected leg second and without help from arms

1 = can not step up 6 inches
2 = 6 inches high
3 = 8 inches high
4 = 10 inches high
5 = 12 inches high

Coronal Tibiofemoral Angle for Valgus

Coronal Tibiofemoral Angle for Varus

Injured Knee

Range of movement

Flexion

Flexion Contracture

Extension Lag

Abduction hip

Foot RVF

Other Knee

Range of movement

Flexion

Flexion Contracture

Extension Lag

Abduction hip

Foot RVF

State of patella
- 1 = Normal function
- 2 = Palella-femoral pain
- 3 = Excised

Patient passed away
- 0 = No
- 1 = Yes

Review Year

Knee Replacement System

Post-Operative Assessment (examples of the entries made)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	9999	0	-------	0	Always
2	Date of Ass.	D	-----	------	------	-------	< (before)	Always
3	Weight	N	2	210	7	-------	70	Always
4	Freq. Pain	N	0	5	1	-------	5	Always
5	Sev. Pain	N	0	5	1	-------	5	Always
6	Pain at night	N	0	5	1	-------	5	Always
7	Walking ability	N	0	5	1	-------	5	Always
8	Walking Aid	N	0	5	1	-------	5	Always
9	Sitting Down	N	0	5	1	-------	5	Always
10	Rising up	N	0	5	1	-------	5	Always
11	Ability to stand	N	0	5	1	-------	5	Always
12	Going upstairs	N	0	5	1	-------	5	Always
13	Going downstairs	N	0	5	1	-------	5	Always
14	Cor. Tib. An. Val	N	2	60	0	-------	0	Always
15	Cor. Tib. An Var	N	2	60	0	-------	0	Always
16	**Injured Knee**	**C**	------	------	------	-------	------	**<Next question>**
16	Rang. movement	N	2	140	0	-------	0	Always
17	Flexion	N	2	140	0	-------	0	Always
18	Flexion Con.	N	2	90	0	-------	0	Always
19	Extension lag	N	2	180	0	-------	0	Always
20	Abduction Hip	N	2	60	0	-------	0	Always
21	Foot RVF	N	2	99	0	-------	0	Always
22	**Other Knee**	**C**	------	------	------	-------	------	**<Next question>**
22	Rang. movement	N	2	140	0	-------	0	Always
23	Flexion	N	2	140	0	-------	0	Always
24	Flexion Con.	N	2	90	0	-------	0	Always
25	Extension lag	N	2	180	0	-------	0	Always
26	Abduction Hip	N	2	60	0	-------	0	Always
27	Foot RVF	N	2	99	0	-------	0	Always
28	State Patella	N	0	3	1	-------	1	Always
29	Pat. passed away	N	0	1	0	-------	0	Always
30	Review Year	N	0	10	0	------	0	Always

The following pages are copies from the report 'Specification of Statistical Analysis for Knee Replacements', submitted by D.K. Bhatt in 1990 as her final year project for the B.Sc. (Hons) Combined Studies course at Leicester Polytechnic. They demonstrate the type and the layout of the reports the medical consultants at Glenfield Hospital, Leicester, can currently obtain from her Knee Replacement system.

```
GET /FILE 'TKR1.SYS'.
The SPSS/PC+ system file is read from
    file TKR1.SYS
The file was created on 9/20/89 at 21:45:40
and is titled SPSS?PC+ System File Written by Data Entry II
The SPSS/PC+ system file contains
    109 cases, each consisting of
    138 variables (including system variables).
    138 variables will be used in this session.
--------------------------------------------------------------
Page    4                          SPSS/PC+

This procedure was completed at 21:54:05

--------------------------------------------------------------
Page    5                          SPSS/PC+

PROCESS IF (Year1 EQ 1).
--------------------------------------------------------------
Page    6                          SPSS/PC+

FREQUENCIES AGE DIAG OP SIDE WEIGHT PAINS PAINN STAR WABP WAP SGUP
GDP CTFVAL CTFVAR FLEXPRE PCPRE ROMPRE EXLAGP PATP WEIGHT1 WEIGHT1Y
PAINS1Y PAINN1Y STAB1Y WAB1Y WA1Y SD1Y RU1Y GU1Y DTFVAL1Y CTFVA1Y
FLEX1Y ROM1Y EXLAG1Y HIPAB1Y FRVF1Y PAT1Y OKF1Y OKFC1Y OKROM1Y
OHAB1Y
/BARCHART /STATISTICS ALL.

***** Memory allows a total of  11836 Values, accumulated across all
      There also may be up to    1479 Value Labels for each Variable

------------------------------------------------------------------
Page    7                          SPSS/PC+

AGE        age at operation
```

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	59	6	11.3	11.3	11.3
	62	1	1.9	1.9	13.2
	63	3	5.7	5.7	18.9
	64	2	3.8	3.8	22.7
	65	2	3.8	3.8	26.5
	66	2	3.8	3.8	30.3
	67	2	3.8	3.8	34.1
	68	1	1.9	1.9	36.0
	69	3	5.7	5.7	41.7
	70	2	3.8	3.8	45.5
	71	3	5.7	5.7	51.2
	72	1	1.9	1.9	53.1
	73	3	5.7	5.7	58.8
	74	3	5.7	5.7	64.5
	75	4	7.5	7.5	72.0
	76	5	9.4	9.4	81.4
	77	2	3.8	3.8	85.2
	78	2	3.8	3.8	89.0

Page 8 SPSS/PC+

Age age at operation

79	2	3.8	3.8	92.8
80	1	1.9	1.9	94.7
81	1	1.9	1.9	96.6
82	1	1.9	1.9	98.5
87	1	1.9	1.9	100.4

Page 9 SPSS/PC+

Age age at operation

```
59 ------------------------------ 6
62 ------ 1
63 ---------------- 3
64 ----------- 2
65 ----------- 2
66 ----------- 2
67 ----------- 2
68 ------ 1
69 ---------------- 3
70 ----------- 2
71 ---------------- 3
72 ------ 1
73 ---------------- 3
74 ---------------- 3
75 --------------------- 4
76 --------------------------5
77 ----------- 2
78 ----------- 2
79 ----------- 2
80 ------ 1
81 ------ 1
```

Page 10 SPSS/PC+

AGE age at operation

```
82 ------ 1
83 ------ 1
```

Mean	70.717	Std Err	.951	Median	71.000
Mode	59.000	Std Dev	6.924	Variance	47.938
Kurtosis	-.682	S E Kurt	.644	Skewness	-.107
S E Skew	.327	Range	28.000	Minimum	59.000
Maximum	87.000	Sum	3748.000		

Valid Cases 53 Missing Cases 0

DIAG diagnosis

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
osteoarthritis	1	45	84.9	84.9	84.9
rheumatoid arthritis	2	2	15.1	15.1	100.0
	TOTAL	53	100.0	100.0	

```
      osteoarthritis ---------------------------------------------- 45
rheumatoid arthritis --------- 8
```

Mean	1.151	Std Err	.050	Median	1.000
Mode	1.000	Std Dev	.361	Variance	.131
Kurtosis	2.108	S E Kurt	.644	Skewness	2.007
S E Skew	.327	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	61.000		

Valid Cases 53 Missing Cases 0

OP type of operation

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
cemented	1	27	50.9	55.1	55.1
uncemented	2	20	37.7	40.8	95.9
uncemented femur,cem	4	1	1.9	2.0	98.0
other	5	1	1.9	2.0	100.0
		4	7.5	MISSING	
	TOTAL	53	100.0	100.0	

```
            cemented ---------------------------------------------- 27
          uncemented ---------------------------------- 20
uncemented femur,cem --- 1
               other --- 1
```

OP type of operation

Mean	1.151	Std Err	.113	Median	1.000
Mode	1.000	Std Dev	.792	Variance	.628
Kurtosis	7.640	S E Kurt	.668	Skewness	2.316
S E Skew	.340	Range	4.000	Minimum	1.000
Maximum	5.000	Sum	76.000		

Valid Cases 49 Missing Cases 4

SIDE operation side

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
right	1	29	54.7	54.7	54.7
left	2	24	45.3	45.3	100.0
	TOTAL	53	100.0	100.0	

```
right ------------------------------------------ 29
 left ---------------------------------- 24
```

Mean	1.453	Std Err	.069	Median	1.000
Mode	1.000	Std Dev	.503	Variance	.253
Kurtosis	-2.040	S E Kurt	.644	Skewness	.195
S E Skew	.327	Range	1.000	Minimum	1.000
Maximum	2.000	Sum	77.000		

Valid Cases 53 Missing Cases 0

WEIGHT WEIGHT PREOP

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	38	1	1.9	2.2	2.2
	44	1	1.9	2.2	4.3
	50	1	1.9	2.2	6.5
	53	1	1.9	2.2	8.7
	58	1	1.9	2.2	10.9
	59	2	3.8	4.3	15.2
	60	1	1.9	2.2	17.4
	61	3	5.7	6.5	23.9
	63	1	1.9	2.2	26.1
	64	2	3.8	4.3	30.6
	65	2	3.8	4.3	34.9
	66	1	1.9	2.2	37.1
	69	1	1.9	2.2	39.3
	70	2	3.8	4.3	43.6
	71	3	5.7	6.5	50.1
	72	3	5.7	6.5	56.6
	73	1	1.9	2.2	58.8
	74	3	5.7	6.5	65.3
	75	2	3.8	4.3	69.6
	76	2	3.8	4.3	73.9
	77	3	5.7	6.5	80.4

```
WEIGHT             WEIGHT PREOP
                             78         1      1.9      2.2     82.6
                             79         1      1.9      2.2     84.8
                             80         1      1.9      2.2     87.0
                             81         1      1.9      2.2     89.1
                             82         1      1.9      2.2     91.3
                             86         1      1.9      2.2     93.5
                             89         1      1.9      2.2     95.7
                             94         1      1.9      2.2     97.8
                             99         1      1.9      2.2    100.0
                                         7     13.2  MISSING
                                    ------  -------  -------
                          TOTAL         53    100.0    100.0
----------------------------------------------------------------
```



```
WEIGHT             WEIGHT PREOP

                     38 ------ 1
                     44 ------ 1
                     50 ------ 1
                     53 ------ 1
                     58 ------ 1
                     59 ---------------- 2
                     60 ------ 1
                     61 ------------------------ 3
                     63 ------ 1
                     64 ---------------- 2
                     65 ---------------- 2
                     66 ------ 1
                     69 ------ 1
                     70 ---------------- 2
                     71 ------------------------ 3
                     72 ------------------------ 3
                     73 ------ 1
                     74 ------------------------ 3
                     75 ---------------- 2
                     76 ---------------- 2
                     77 ------------------------ 3
                     78 ------ 1
                     79 ------ 1
                     80 ------ 1
                     81 ------ 1
                     82 ------ 1
                     86 ------ 1
                     89 ------ 1
                     94 ------ 1
                     99 ------ 1
----------------------------------------------------------------
```

Page 18 SPSS/PC+

WEIGHT WEIGHT PREOP

Mean	70.109	Std Err	1.729	Median	71.500
Mode	61.000	Std Dev	11.729	Variance	137.566
Kurtosis	1.010	S E Kurt	.688	Skewness	-.252
S E Skew	.350	Range	61.000	Minimum	38.000
Maximum	89.000	Sum	3225.000		

Valid Cases 46 Missing Cases 7

Page 19 SPSS/PC+

PAINF FREQUENCY OF PAIN PREOP

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	6	11.3	11.3	11.3
	2	35	66.0	66.0	77.4
	3	10	18.9	18.9	96.2
	4	2	3.8	3.8	100.0
	TOTAL	53	100.0	100.0	

```
1 -------- 6
2 --------------------------------------------------------- 35
3 ------------- 10
4 --- 2
```

Page 20 SPSS/PC+

PAINF FREQUENCY OF PAIN PREOP

Mean	2.151	Std Err	.091	Median	2.000
Mode	2.000	Std Dev	.662	Variance	.438
Kurtosis	1.214	S E Kurt	.644	Skewness	.654
S E Skew	.327	Range	3.000	Minimum	1.000
Maximum	4.000	Sum	114.000		

Valid Cases 53 Missing Cases 0

Page 21 SPSS/PC+

PAINS SEVERITY OF PAIN PREOP

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	1	1.9	1.9	1.9
	2	32	60.4	60.4	62.3
	3	18	34.0	34.0	96.2
	4	2	3.8	3.8	100.0
	TOTAL	53	100.0	100.0	

Page 23 SPSS/PC+

PAINN PAIN AT NIGHT PREOP

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	8	15.1	15.1	15.1
	2	11	20.8	20.8	35.8
	3	13	24.5	24.5	60.4
	4	21	39.6	39.6	100.0
	TOTAL	53	100.0	100.0	

```
1 ----------------- 8
2 ----------------------- 11
3 --------------------------- 13
4 ------------------------------------------- 21
```

Page 24 SPSS/PC+

PAINN PAIN AT NIGHT PREOP

Mean	2.889	Std Err	.152	Median	3.000
Mode	4.000	Std Dev	1.103	Variance	1.218
Kurtosis	-1.137	S E Kurt	.644	Skewness	-.482
S E Skew	.327	Range	3.000	Minimum	1.000
Maximum	4.000	Sum	153.000		

Valid Cases 53 Missing Cases 0

Page 25 SPSS/PC+

STAB STABILITY PREOP

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	6	11.3	11.3	11.3
	2	16	30.2	30.2	41.5
	3	20	37.7	37.7	79.2
	4	8	15.1	15.1	94.3
	5	3	5.7	5.7	100.0
	TOTAL	53	100.0	100.0	

```
1 -------------- 6
2 -------------------------------- 16
3 -------------------------------------------------- 20
4 ------------------- 8
5 -------- 3
```

Appendix C (VIII)

Patella-femoral trial (III) and comparing two Knee Ligament systems and specifying the benefits of such systems

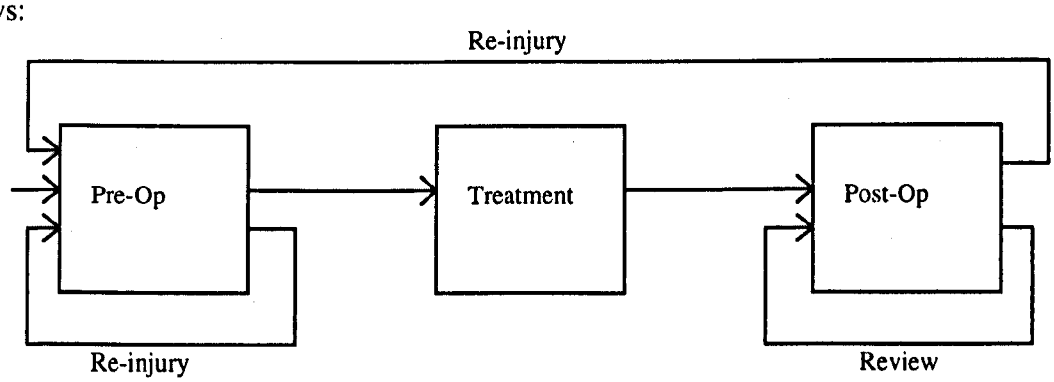
The reason for this trial was three fold: firstly, to determine if a medical consultant, with very limited experience of computers, could successfully use the tools to build his own IDDA end-system; secondly, for the medical consultant to compare a knee ligament system that was generated by the tools with a system that had been hand-built; and thirdly, to obtain from the consultant his views regarding the benefits that such systems could bring to clinical care.

This trial was undertaken by an orthopaedic consultant surgeon, Mr Harding, who has very limited experience of operating a computer (the previous 6 months word-processing on a PC at home).

a) Constructing a Patella-femoral system using the tools:

The first section of this trial was therefore to determine if Mr Harding could successfully build an IDDA end-system for a Patella-femoral study.

Mr Harding had rewritten existing manual Patella-femoral questionnaires into three assessments: a Pre-operative assessment consisting of 30 questions, a Treatment assessment of 9 questions and a Post-operative assessment of 28 questions. The original questionnaire (see Appendix C (IV)) had taken approximately 2 hours to compose, the rewritting took less than 10 minutes. In addition to the assessment questionnaires, Mr Harding also produced small sections of help text for 3 of the assessment questions. For this trial, the assessments and the help text had already been entered into a word-processor. The assessment process for a patient suffering from patella femoral pain would be as follows:



Points to note:

Mr Harding successfully built the required IDDA system. The time taken to produce the end-system was just 1 hour, even though no help documentation explaining the tools had been made available to

Mr Harding prior to the evaluation. There was, however, on-line help freely accessible during the trial.

Mr Harding reported that he found the menus, instructions, diagrams and error messages easy to understand and follow. He thought the Help was easy to use as it required only one keystroke and was always associated with that one key. However, other than experimenting with the facility, he did not reference the help during the trial.

With regards to the general screen layout, Mr Harding recorded that he thought that sufficient information was given, without causing the screen to appear too cluttered and that the error messages were easy to locate and read. He was also happy with the current selection of screen colours.

The role of the second screen did not confuse or irritate him. In fact, he said he preferred referencing another source, i.e. a book or other screen, rather than having the help displayed over the original information. In response to the query regarding the space that the second monitor would require, Mr Harding responded by saying he did not perceive that a problem would exist as it could be placed on a separate table in an L-shaped layout in the corner of the room. In addition, the assistance that the second terminal could give throughout the analysis section made it, in his opinion, an important component of the end-system.

It does seem that the impact of the second screen will only be determined after extended use of the end-system. This is evident from the conflicting remarks made during the evaluations that have been carried out so far: some people like the second screen being present, others do not, whilst some would like the ability to turn it off and on depending upon where the end-system is to be operated. Therefore the value of the second monitor will have to be assessed and monitored over time, use and within different situations before any conclusions regarding its role can be drawn.

The speed of the computer, Mr Harding felt, was fine, neither too fast nor too slow and, finally, he thought that the responses and actions of the computer system were consistent. Therefore Mr Harding found that, overall, the tools were easy to operate and he quickly felt confident whilst using them.

The only difficulty, he reported was that of defining when a question should be asked (always or only when). This it seems caused a little confusion the first time a condition needed to be specified for a question, i.e. I want to only ask this question only when - *such and such ocurred.* However, after completing this operation once, Mr Harding seemed much happier the next time he was required to specify a condition. It must be noted, that prior to the trial, Mr Harding had not been given any written instructions on how to use the tools, only a brief verbal introduction and an explanation of certain functions. One part of this introduction had been a very brief explanation of the facility which defined when a question had to be asked. However, the first time when this facility was to be used was over half way through the trial. It is therefore reasonable to expect some difficulty during the first attempt to specify a condition. The fact that Mr Harding seemed happier as he defined other conditions later in the trial and, as he stated, he 'did not think any additions or alterations were needed to the condition section as he was now aware of what to expect', seems to indicate that for this section there must be a more detailed explanation with examples at the beginning of the very first trial. If this is done, it would appear that users will be able to carry out the required operations with increasing confidence and ease. In other words, it seems that the condition section requires practice, although it can be successfully completed without any prior knowledge.

Mr Harding reported no other difficulties in operating the tools and could suggest no further improvements. He was impressed with the progress of the trial and was surprised at how easy he had found the tools to operate and thus how easy it had been to produce his own end-system.

b) Comparing the Knee Ligament system generated by the tools with one that was hand-built.

When comparing the IDDA end-system generated by the tools (see pages C-73 to C-88) with the hand-built system, Mr Harding reported that the generated end-system could carry out the required job of data collection as easily as the hand-built system. In addition, he thought the generated IDDA system provided greatly improved methods of access to the information and better techniques to alter the stored data. Moreover, he viewed the extra facilities within the new system for analysing data and displaying results in various forms as being very useful and a very important addition.

Mr Harding reported that the analysis facilities offered by the IDDA system reflected his original idea of the methods consultants would use to inspect their data stored on a computer, 'just put the data in and then get it out [by selecting it] and manipulate it in anyway you like. In other words, declare what factor you want to pick out and analyse it against whatever you like and then display the result.' The ability to display the relevant study numbers involved in each of the categories in a graph was also thought to be very useful as it allowed cross-checking by hand, if the consultant deemed it necessary. Consequently, Mr Harding thought the generated system had distinct advantages over the hand-built version, especially in the areas of data analysis and reviews. Furthermore, the ease of operation of the end-system meant that Mr Harding was confident with secretarial support staff carrying out the data entry, thereby releasing him to undertake further investigations.

The only feature missing from the generated IDDA system was a printing facility. However, this would be an easy addition by:

a) including another facility to re-direct reviews to the printer rather than the screen,
b) including relevant messages on the screen to prompt the user when screen dumps would be appropriate.

With regards to actually building his own system, Mr Harding thought that this ability was highly beneficial as it gives consultants the necessary control over the construction of their end-systems and that they could therefore tailor it to the working practices of the medical unit. In addition, it would make the whole process 'an awful lot faster if the doctor could build his own system.' This is a very important ability as medicine evolves so quickly and new studies are therefore being intiated continually.

The following example given by Mr Harding demonstrates this trait. Another international knee scoring system has just been proposed that is different yet again from the old Lysholm, the Activity or the British knee scoring systems. These older tests are still being used but consultants are being asked to collect any new information by following the new scheme. There has been no comparative study of these various scoring systems and so consultants are rather reluctant to keep switching. Especially, as Mr Harding pointed out, because of the requirement that the Post-operative assessment techniques have to be similar to the Pre-operative methods to enable comparisons and analyses to be undertaken to determine the progression of a patient. Therefore, if a patient has already been assessed using one of the old scoring systems, that patient must continue to be assessed by that method. This results in various patient groups following different assessment paths. This makes consultants wary of volunteering to take part in assessing yet another technique when no evaluation of the current or previous methods have been undertaken.

For those doctors who wish to try the new method, a new study will have to be initiated and thus a new computer system will have to be built to collect the appropriate details. Within this kind of dynamic environment, the ability to quickly construct an appropriate computer system oneself has obvious benefits since it enables the advantages of computerisation to be realised quickly, easily and with very little disruption to the patients or the specialist unit.

c) The benefits that such systems could bring to clinical care

This third section of the evaluation was concerned with Mr Harding identifying, in his opinion, the benefits computer systems similar to the IDDA end-system could bring to the specialist medical field. He responded by saying that he believes clinical care would be improved by using such systems as the IDDA end-system, although he thinks that they might interfere with the patient-physician relationship by requiring extra sessions for data entry. Even though he had these reservations, he stated that such systems would fit in easily to the daily working practices of his unit and would not delay the work unduly. When it was explained that a manual recording system could be used during patient assessment sessions, he considered that to be a better approach to adopt. It would also allow the assessment details to be entered into the system by the secretarial support staff, thus speeding up that stage of the process. He could see no reason why the present secretaries could not operate the end-system, if, as he put it, 'he could' since they all had more experience of using a computer than himself.

The main purpose of such systems in specialist medical fields, Mr Harding believed, was to ensure the thorough and consistent collection of data and to assist in reviews and analyses of this data. Having seen the IDDA end-system, he could not identify at this time any improvements that should be made to either the IDDA tools or the end-system as he had not used them long enough to determine their limitations. He did state however that he was impressed with the current facilities that were offered.

In fact, Mr Harding has requested to use the IDDA tools to build his own end-system for another study he is about to initiate. He has been awarded a grant by The Welcome Foundation to undertake an investigation into knee injuries and is currently in the process of purchasing the computer hardware and software required for the operation of the IDDA tools. As soon as the equipment is delivered and installed, Mr Harding will use the IDDA tools to construct for himself the appropriate end-system for this study. He will then utilise the end-system to collect, review and analyse his data. Hence, this new study will enable further on-going evaluations to be undertaken regarding the tools, the end-system and their impact on the medical field and environment, thus enabling the study of the longer-term effects of the tools and the IDDA system.

Knee Ligament System

Pre-operative Assessment

STUDY NUMBER

DATE OF ASSESSMENT

SEX
1. Male
2. Female

DATE OF BIRTH

DATE OF INJURY

RE-INJURY
0. No
1. Yes

DATE OF PRE-OPERATIVE ASSESSMENT

By Study Team (or by M.L.H. alone). If there is doubt, record the date which is specific to the definitive ligament reconstruction, i.e. the date seen and put on waiting list for example.

AETIOLOGY
0. Not known
1. Sport 2. Car 3. Motorbike
4. Industria 5. Domestic 6. Other

SIDE
1. Right
2. Left

ASSESSMENT OF OTHER JOINT
0. Not known
1. Problem but not assessed
2. Problem and assessed
3. Other knee normal

PREVIOUS INJURY TO JOINT
0. Not known
1. No injury
2. Yes

PREVIOUS TREATMENT
0. Not known
1. None
2. General Practitioner
3. Accident and Emergency
4. Orthopaedic
5. Non-specific - physio. etc.

ABILITY TO WORK AT TIME OF PRE-OPERATIVE ASSESSMENT
0. Not known
1. No
2. Yes
3. Not applicable

POPLITEAL NERVE INJURY

0. Not known
1. Absent, no palsy
2. Present, palsy

SPORTING ACTIVITY LEVEL

0. Not known
1. Knee gives way when walking
2. Cannot run
3. Cannot play tennis, squash or equivalent.
4. Cannot play football
5. Normal

LYSHOLM ASSESSMENT

LIMP

5. None
3. Slight or Periodical
0. Severe or Constant

SUPPORT

5. None
3. Stick or Crutch
0. Weight Bearing Impossible

LOCKING

15. No Locking or Catching Sensation
10. Catching Sensation but No Locking
6. Locking Occasionally
2. Locking Frequently
0. Locked Joint on Examination

INSTABILITY

25. Never Giving Way
20. Rarely During Athletics or Similar
15. Frequently during Athletics or Similar
10. Occasionally in Daily Activity
5. Often in Daily Activity
0. Every Step

PAIN

25. None
20. Inconstant and Slight During Severe Exertion
15. Marked During Severe Exertion
10. Marked on or after Walking > 2km
5. Marked on or after Walking < 2km
0. Constant

SWELLING

10. None
6. On Severe Exertion
2. On Ordinary Exertion
0. Constant

STAIR CLIMBING

10. No Problems
6. Slightly Impaired
2. One Step at a Time
0. Impossible

SQUATTING
5. No Problem
4. Slightly Impaired
2. Not Beyond 90
0. Impossible

RECORD FINAL SCORE
(Worse is 0. Best is 100)

ARE X-RAYS AVAILABLE
0. Not known
1. No
2. Yes

VALUE ?
0. Not Known
1. Normal
2. Abnormal

AVULSION FRACTURE(s) FROM COLLATERAL LIGAMENTS:
0. Not known
1. No injury
2. Fractures present

AVULSION FRACTURE(S) FROM CRUCIATES:
0. Not known
1. No avulsion
2. Fractures present

INTRA-ARTICULAR FRACTURE(S) (NOT AVULSION TYPE)
0. Not Known
1. No fractures
2. Fractures present

DISLOCATION OF PATELLA
0. Not known
1. No dislocation
2. Dislocation present

FRACTURE OF PATELLA
0. Not known
1. No fracture
2. Fracture present

DISLOCATION OF KNEE
0. Not known
1. No dislocation
2. Dislocation occurred

FRACTURE OF FEMORAL SHAFT
0. Not known
1. No fracture
2. Fracture present

FRACTURE OF FEMORAL CONDYLE(S)
0. Not known
1. No fracture
2. Fracture present

FRACTURE OF TIBIAL SHAFT
0. Not known
1. No fracture
2. Fracture present

FRACTURE OF TIBIAL PLATEAU(X)
0. Not known
1. No fracture
2. Fracture present

OTHER(S) (Specify)
OLD CHANGES
0. Not known
1. None
2. Present

PIVOT SHIFT UNDER G.A.
0. Not known
1. Negative
2. Positive
3. Equivocal

CLINICAL ASSESSMENT OF SAGITTAL LAXITY
0. Not known
1. None
2. +/- = slight
3. + = to 1 cm
4. ++ = 1-2 cm
5. +++ = more than 2 cm or gross

WRITE IN MEASURED ANTERIOR LAXITY IN MMs at 0

WRITE IN MEASURED ANTERIOR LAXITY IN MMs at 20

WRITE IN MEASURED ANTERIOR LAXITY IN MMs at 90

WRITE IN MEASURED ANTERIOR LAXITY OF NORMAL KNEE IN MMs at 0

WRITE IN MEASURED ANTERIOR LAXITY OF NORMAL KNEE IN MMs at 20

WRITE IN MEASURED ANTERIOR LAXITY OF NORMAL KNEE IN MMs at 90

WRITE IN ANTERIOR STIFFNESS OF INJURED KNEE at 20

WRITE IN ANTERIOR STIFFNESS OF NORMAL KNEE at 20

WRITE IN MEASURED POSTERIOR LAXITY IN MMs at 0

WRITE IN MEASURED POSTERIOR LAXITY IN MMs at 20

WRITE IN MEASURED POSTERIOR LAXITY IN MMs at 90

WRITE IN MEASURED POSTERIOR LAXITY OF NORMAL KNEE IN MMs at 0

WRITE IN MEASURED POSTERIOR LAXITY OF NORMAL KNEE IN MMs at 20

WRITE IN MEASURED POSTERIOR LAXITY OF NORMAL KNEE IN MMs at 90

WRITE IN POSTERIOR STIFFNESS OF INJURED KNEE at 90

WRITE IN POSTERIOR STIFFNESS OF NORMAL KNEE at 90

Knee Ligament System

Pre-Operative Assessment (examples of the entries made)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	999999	0	-------	0	Always
2	Date of Ass.	D	-----	------	------	-------	< (before)	Always
3	Sex of Patient	N	0	2	1	-------	1	Always
4	Date of Birth	D	-----	------	------	-------	< (before)	Always
5	Date of injury	D	-----	------	------	-------	< (before)	Always
6	Re-injury	N	0	1	0	------	0	Always
7	Date of pre-op	D	-----	------	------	-------	< (before)	Always
8	**By study team**	**C**	-----	------	------	------	------	**<Prev. question>**
8	Aetiology	N	0	6	0	------	0	Always
9	Side	N	0	2	1	------	1	Always
10	Ass. Other joint	N	0	3	0	------	0	Always
11	Prev injury	N	0	2	0	------	0	Always
12	Prev. treatment	N	0	5	0	-------	0	Always
13	Ability to work	N	0	3	0	-------	0	Always
14	Pop. nerve injury	N	0	2	0	-------	0	Always
15	Sporting level	N	0	5	0	-------	0	Always
16	**Lysholm**	**C**	------	------	------	-------	------	**<Next question>**
16	Limp	N	0	5	0	-------	5	Always
17	Support	N	0	5	0	-------	5	Always
18	Locking	N	0	15	0	-------	15	Always
19	Instability	N	0	25	0	-------	25	Always
20	Pain	N	0	25	0	-------	25	Always
21	Swelling	N	0	10	0	-------	10	Always
22	Stair climbing	N	0	10	0	-------	10	Always
23	Squatting	N	0	5	0	-------	5	Always
24	Final score	N	0	100	0	-------	100	Always
25	X-rays	N	0	2	0	-------	0	Always
26	Value	N	0	2	0	-------	0	Always
27	Av. F. Coll.	N	0	2	0	-------	0	Always
28	Av. F. Cr	N	0	2	0	-------	0	Always
29	Intra-art. frac.	N	0	2	0	-------	0	Always
30	Disloc. pat.	N	0	2	0	-------	0	Always
31	Frac. pat.	N	0	2	0	-------	0	Always
32	Disloc. knee	N	0	2	0	-------	0	Always
33	Frac. fem. shaft	N	0	2	0	-------	0	Always
34	Frac. fem. condy.	N	0	2	0	-------	0	Always
35	Frac. tib. shaft	N	0	2	0	-------	0	Always
36	Frac. tib. plateaux	N	0	2	0	-------	0	Always
37	Other	A	------	------	------	20	<blank>	Always
38	Old changes	N	0	2	0	-------	0	Always
39	Pivot shaft	N	0	3	0	-------	0	Always
40	Clin. sag. lax.	N	0	5	0	-------	0	Always
41	Ant. 0	N	2	99	0	-------	99	Always
42	Ant. 20	N	2	99	0	-------	99	Always
43	Ant. 90	N	2	99	0	-------	99	Always
44	Ant. other 0	N	2	99	0	-------	99	Always
45	Ant. other 20	N	2	99	0	-------	99	Always
46	Ant. other 90	N	2	99	0	-------	99	Always
47	Ant. Stiff. inj. 20	N	2	99	0	-------	99	Always
48	Ant. Stiff. oth. 20	N	2	99	0	-------	99	Always

49	Pos. 0	N	2	99	0	-------	99	Always
50	Pos 20	N	2	99	0	-------	99	Always
51	Pos. 90	N	2	99	0	-------	99	Always
52	Pos. other 0	N	2	99	0	-------	99	Always
53	Pos. other 20	N	2	99	0	-------	99	Always
54	Pos. other 90	N	2	99	0	-------	99	Always
55	Pos. stiff. inj. 90	N	2	99	0	-------	99	Always
56	Pos. stiff. oth. 90	N	2	99	0	-------	99	Always

Knee Ligament System

Treatment Assessment

STUDY NUMBER

DATE OF ASSESSMENT

DATE PREVIOUS SURGERY OR TREATMENT

TREATMENT (Highlight most important)

0. Not known
1. None
2. Strapping
3. Strapping and physiotherapy
4. Plaster cast only
5. Plaster cast and physiotherapy
6. Knee brace
7. Cast brace
8. Physiotherapy alone
9. Surgery

DATE OF DEFINITIVE SURGERY: ....

PREVIOUS SURGERY TO SAME JOINT
(Count arthroscopy alone as a non-surgical event)

0. Not known
1. No previous surgery
2. Yes

LIGAMENTS INJURED

ACL
0. Not known
1. No ligament tear
2. Partial
3. Complete*

* Record complete if at least one ligament is disrupted.

PCL
0. Not known
1. No tear
2. Partial
3. Complete

LCL AND LATERAL COMPLEX (Popliteus, arcuate ligament)

0. Not known
1. No tear
2. Partial
3. Complete

MCL AND MEDIAL COMPLEX EXCLUDING POSTERIOR OBLIQUE LIGAMENT

0. Not known
1. No tear
2. Partial
3. Complete

POSTERIOR OBLIQUE LIGAMENT

0. Not known
1. No tear
2. Partial
3. Complete

POSTERIOR CAPSULE

0. Not known
1. No tear
2. Partial
3. Complete

MEDIAL MENISCUS

0. Not known
1. No tear
2. Partial
3. Complete

LATERAL MENISCUS

0. Not known
1. No tear
2. Partial
3. Complete

CHONDRAL SURFACE DAMAGE

a) PATELLA

0. Not known
1. Normal
2. Slight
3. Moderate
4. Full thickness

b) MFC

0. Not known
1. Normal
2. Slight
3. Moderate
4. Full thickness

c) MTP

0. Not known
1. Normal
2. Slight
3. Moderate
4. Full thickness

d) LFC

0. Not known
1. Normal
2. Slight
3. Moderate
4. Full thickness

e) LTP

0. Not known
1. Normal
2. Slight
3. Moderate
4. Full thickness

NOTE

Slight	2	= Meachim Grade I
Moderate	3	= Meachim Grade II
Full thickness	4	= Meachim Grade III

or seven

SURGICAL PROCEDURE

ACL 0. Not known 1. None 2. Primary repair
3. Reconstruction with patella tendon (Jones)
4. Reconstruction with patella tendon (MacIntosh)
5. Reconstruction with fascialata (inc. MacIntosh III)
6. Reconstruction with Dacron
7. Reconstruction with Surgicraft ABC ligament
8. Reconstruction with carbon fibre
9. Primary Repair Augmented with Dacron
10. Primary Repair Augmented with ABC
11. Primary Repair Augmented with Carbon Fibre
12. Other

ACL Other ....

PCL
0. Not known 1. None
2. Primary repair
3. Reconstruction with tendon/ligament
4. Reconstruction with carbon fibre
5. Reconstruction with Dacron
6. Reconstruction with ABC ligament
7. Primary Repair augmented with Carbon Fibre
8. Primary Repair augmented with Dacron
9. Primary Repair augmented with ABC
10. Primary Repair augmented with Dexon

LCL (NOTE: This is NOT the section for lateral stablising procedure for ACL insufficiency - see question 58.)
0. Not known 1. None 2. Primary repair
3. Reconstruction with fascia lata
4. Reconstruction with carbon fibre
5. Reconstruction with Dacron
6. Reconstruction with ABC ligament
7. Primary Repair augmented with Carbon Fibre
8. Primary Repair augmented with Dacron
9. Primary Repair augmented with ABC
10. Primary Repair augmented with Dexon

LATERAL EXTRA ARTICULAR STABLISING PROCEDURE FOR ACL INSUFFICIENCY
0. Not known
1. None
2. Reconstruction with fascia lata (MacIntosh III type)
3. Reconstruction with fascia lata (LaMaire type)
4. Reconstruction with carbon fibre
5. Reconstruction with Dacron
6. Reconstruction with ABC ligament

LATERAL COMPLEX STRUCTURES OTHER THAN LCL, I.E. POPLITEUS TENDON ARCUATE LIGAMENT POSTERO LATERAL CAPSULE
0. Not known 1. None 2. Primary repair
3. Reconstruction with fascia lata
4. Reconstruction with carbon fibre
5. Reconstruction with Dacron
6. Reconstruction with ABC ligament
7. Primary repair augmented with Carbon Fibre
8. Primary Repair augmented with Dacron
9. Primary Repair augmented with ABC
10. Primary Repair augmented with Dexon

MCL SUPERFICIAL AND OR DEEP. NOT POSTERIOR OBLIQUE LIGAMENT ALONE

0. Not known 1. None
2. Primary repair
3. Reconstruction with proximal realignment
4. Reconstruction with distal realignmenmt
5. Reconstruction with reefing
6. Carbon fibre recon. 7. Dacron recon.
8. ABC reconstruction
9. Pri. repair Car. Fibre 10. Pri. Repair Dacron
11. Pri. Repair ABC 12. Pri. Repair Dexon
13. Other

MCL OTHER ....

MCL POSTERIOR OBLIQUE LIGAMENT OR POSTERO MEDIAL CORNER RECONSTRUCTION. APPROPRIATE PART OF 5 IN 1 PROCEDURE (NICHOLAS) SHOULD BE RECORDED

0. Not known 1. None
2. Primary repair
3. Reconstruction with reefing or overlap (i.e. Nicholas component)
4. Carbon fibre reconstruction
5. Dacron reconstruction
6. ABC reconstruction
7. Pri. repair Car. Fibre 8. Pri. Repair Dacron
9. Pri. Repair ABC 10. Pri. Repair Dexon

POSTERIOR CAPSULE

0. Not known 1. None
2. Primary repair
3. Reconstruction with reefing or overlap
4. Carbon fibre reconstruction
5. Dacron reconstruction
6. ABC reconstruction
7. Primary repair augmented with Carbon Fibre
8. Primary Repair augmented with Dacron
9. Primary Repair augmented with ABC
10. Primary Repair augmented with Dexon

MEDIAL MENISCUS

0. Not known
1. None
2. Partial meniscectomy
3. Complete meniscectomy
4. Previous meniscectomy (of any sort)
5. Sutured

LATERAL MENISCUS

0. Not known
1. None
2. Partial meniscectomy
3. Complete meniscectomy
4. Previous meniscectomy (any sort)
5. Sutured

PATELLA

0. Not known
1. No procedure
2. Lateral release alone
3. Realignment (Distal)
4. Realignment (Proximal)
5. Chondrectomy + drilling
6. Prosthetic replacement
7. Patellectomy

PRIDIE TYPE (on femoral condyle)

0. Not known
1. No procedure
2. Yes - MCL
3. Yes - LCL
4. Yes Both condyles

PRINCIPAL RECONSTRUCTION MATERIAL

0. Not known
1. None
2. Fascia
3. Ligament or tendon
4. Carbon fibre
5. Dexon
6. Dacron
7. ABC

Knee Ligament System

Treatment Assessment (examples of the entries made)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	999999	0	-------	0	Always
2	Date of Ass.	D	-----	------	------	-------	< (before)	Always
3	Date of Prev. sur.	D	-----	------	------	-------	< (before)	Always
4	Treatment	N	0	9	0	------	0	Always
5	Date of Def. Sur.	D	-----	------	------	-------	< (before)	Always
6	Prev. Surgery	N	0	2	0	------	0	Always
7	**Ligaments Inj.**	**C**	-----	------	------	------	------	**<Next question>**
7	ACL	N	0	3	0	------	0	Always
8	PCL	N	0	3	1	------	1	Always
9	LCL and lat. comp.	N	0	3	0	------	0	Always
10	MCL	N	0	3	0	------	0	Always
11	Pos. ob. lig.	N	0	3	0	-------	0	Always
12	Pos. capsule	N	0	3	0	-------	0	Always
13	Medial Meniscus	N	0	3	0	-------	0	Always
14	Lat. Meniscus	N	0	3	0	-------	0	Always
15	**Chond. Sur. dam**	**C**	------	------	------	-------	------	**<Next question>**
15	Patella	N	0	4	0	-------	0	Always
16	MFC	N	0	4	0	-------	0	Always
17	MTP	N	0	4	0	-------	0	Always
18	LFC	N	0	4	0	-------	0	Always
19	LTP	N	0	4	0	-------	0	Always
20	**NOTE:**	**C**	------	------	------	-------	------	**<Prev. question>**
20	**Sur. Proc.**	**C**	------	------	------	-------	------	**<Next question>**
20	ACL	N	0	12	0	-------	12	Always
21	ACL Other	A	------	------	------	30	<blank>	Q20 = 12
22	PCL	N	0	10	0	-------	0	Always
23	LCL	N	0	10	0	-------	0	Always
24	Lat. extra art.	N	0	6	0	-------	0	Always
25	Lat. complex str.	N	0	10	0	-------	0	Always
26	MCL superfi.	N	0	15	0	-------	0	Always
27	MCL Other	A	------	------	------	30	<blank>	Q26 = 15
28	MCL Pos. ob. lig.	N	0	10	0	-------	0	Always
29	Pos. capsule	N	0	10	0	-------	0	Always
30	Medial Meniscus	N	0	5	0	-------	0	Always
31	Lat. Meniscus	N	0	5	0	-------	0	Always
32	Patella	N	0	7	0	-------	0	Always
33	Pridie type	N	0	4	0	-------	0	Always
34	Prin. Const. Mat.	N	0	7	0	------	0	Always

Knee Ligament System

Post-operative Assessment

STUDY NUMBER

DATE OF FOLLOW UP

SUBJECTIVE ASSESSMENT

0. Not known
1. No change
2. Slight improvement
3. Much improved
4. Normal
5. Worse

LYSHOLM SCORE.

LIMP

5. None.
3. Slight or Periodical
0. Severe and Constant

SUPPORT

5. None
3. Stick or Crutch
0. Weight Bearing Impossible

LOCKING

15. No Locking or Catching Sensation
10. Catching Sensation But No Locking
6. Locking Occasionally
2. Locking Frequently
0. Locked Joint on Examination

INSTABILITY

25. Never Giving Way
20. Rarely during athletics or similar
15. Frequently during athletics or similar
10. Occasionally in Daily Activity
5. Often in Daily Activity
0. Every Step

PAIN

25. None
20. Inconstant and Slight during Severe Exertion
15. Marked during Severe Exertion
10. Marked on or after walking > 2km
5. Marked on or after walking < 2km
0. Constant

SWELLING

10. None
6. On Severe Exertion
2. On Ordinary Exertion
0. Constant

STAIR CLIMBING

10. No Problems
6. Slighty Impaired
2. One Step at a time
0. Impossible

SQUATTING

5. No Problem
4. Slightly Impaired
2. Not beyond 90
0. Impossible

RECORD FINAL SCORE
(Worse is 0. Best is 100)

WORK:

0. Not known
1. Not returned
2. Could return but unemployed
3. Returned to same employment
4. Returned to less demanding employment
5. Returned to more demanding employment

EFFUSION:

0. Not known
1. None
2. Mild
3. Moderate
4. Gross

SPORTING ACTIVITY LEVEL
(Ask "What you can not do ?")

0. Not Known
1. Knee gives way when Walking
2. Cannot Run
3. Cannot play Tennis, Squash or equivalent
4. Cannot play Football
5. Normal

MOVEMENTS (Write in range)

FLEXION DEFORMITY : (Write in flexion deformity)

HYPEREXTENSION POSTURE: (Write in degrees)

TENDERNESS :

0. Not known
1. None
2. Medial joint line
3. Lateral joint line
4. Medial ligament
5. Lateral ligament
6. Retro-patella
7. Non-specific
8. Other

SPECIFY TENDER OTHER ....

PIVOT SHIFT

0. Not known
1. No
2. Yes
3. Equivocal
4. Invalid - will not relax

CLINICAL ASSESSMENT OF SAGITTAL LAXITY:

0. Not known
1. None
2. +/- = slight
3. + = to 1 cm
4. 2+ = 1-2 cm
5. 3+ = more than 2 cm or gross

IS OTHER KNEE NORMAL?

1. No
2. Yes

WRITE IN MEASURED ANTERIOR LAXITY IN MMs at 0

WRITE IN MEASURED ANTERIOR LAXITY IN MMs at 20

WRITE IN MEASURED ANTERIOR LAXITY IN MMs at 90

WRITE IN MEASURED ANTERIOR LAXITY IN MMs OF OTHER KNEE at 0

WRITE IN MEASURED ANTERIOR LAXITY IN MMs OF OTHER KNEE at 20

WRITE IN MEASURED ANTERIOR LAXITY IN MMs OF OTHER KNEE at 90

WRITE IN ANTERIOR STIFFNESS OF INJURED KNEE at 20

WRITE IN ANTERIOR STIFFNESS OF NORMAL KNEE at 20

WRITE IN MEASURED POSTERIOR LAXITY IN MMs at 0

WRITE IN MEASURED POSTERIOR LAXITY IN MMs at 20

WRITE IN MEASURED POSTERIOR LAXITY IN MMs at 90

WRITE IN MEASURED POSTERIOR LAXITY IN MMs OF OTHER KNEE at 0

WRITE IN MEASURED POSTERIOR LAXITY IN MMs OF OTHER KNEE at 20

WRITE IN MEASURED POSTERIOR LAXITY IN MMs OF OTHER KNEE at 90

WRITE IN POSTERIOR STIFFNESS OF INJURED KNEE at 90

WRITE IN POSTERIOR STIFFNESS OF NORMAL KNEE at 90

NAME OF ASSESSOR

Knee Ligament System

Post-Operative Assessment (examples of the entries made)

Question Number	Question	Type	Dec. Places	Max. Answer	Min. Answer	Length	Common Answer	When is the question asked?
1	Patient Study	N	0	999999	0	-------	0	Always
2	Date of Follow up	D	-----	------	------	-------	< (before)	Always
3	Subj. Assess.	N	0	5	0	-------	0	Always
4	**Lysholm**	**C**	------	------	------	-------	------	**<Next question>**
4	Limp	N	0	5	0	-------	5	Always
5	Support	N	0	5	0	-------	5	Always
6	Locking	N	0	15	0	-------	15	Always
7	Instability	N	0	25	0	-------	25	Always
8	Pain	N	0	25	0	-------	25	Always
9	Swelling	N	0	10	0	-------	10	Always
10	Stair climbing	N	0	10	0	-------	10	Always
11	Squatting	N	0	5	0	-------	5	Always
12	Final score	N	0	100	0	-------	100	Always
13	Work	N	0	5	0	-------	0	Always
14	Pop. nerve injury	N	0	4	0	-------	0	Always
15	Sporting level	N	0	5	0	-------	0	Always
16	Movements	N	2	270	0	-------	230	Always
17	Flexion def.	N	2	100	0	-------	10	Always
18	Hyper-ext.	N	2	50	0	-------	0	Always
19	Tenderness	N	0	8	0	-------	0	Always
20	Other	A	------	------	------	20	<blank>	Q19 = 8
21	Pivot shaft	N	0	4	0	-------	0	Always
22	Clin. sag. lax.	N	0	5	0	-------	0	Always
23	Oth. Knee. Nor.	N	0	2	1	-------	2	Always
24	Ant. 0	N	2	99	0	-------	99	Always
25	Ant. 20	N	2	99	0	-------	99	Always
26	Ant. 90	N	2	99	0	-------	99	Always
27	Ant. other 0	N	2	99	0	-------	99	Always
28	Ant. other 20	N	2	99	0	-------	99	Always
29	Ant. other 90	N	2	99	0	-------	99	Always
30	Ant. Stiff. inj. 20	N	2	99	0	-------	99	Always
31	Ant. Stiff. oth. 20	N	2	99	0	-------	99	Always
32	Pos. 0	N	2	99	0	-------	99	Always
33	Pos 20	N	2	99	0	-------	99	Always
34	Pos. 90	N	2	99	0	-------	99	Always
35	Pos. other 0	N	2	99	0	-------	99	Always
36	Pos. other 20	N	2	99	0	-------	99	Always
37	Pos. other 90	N	2	99	0	-------	99	Always
38	Pos. stiff. inj. 90	N	2	99	0	-------	99	Always
39	Pos. stiff. oth. 90	N	2	99	0	-------	99	Always
40	Name Assessor	A	-----	------	------	30	<blank>	Always

Appendix D

Questionnaires:

i) Evaluation Questionnaire

ii) 'Benefits of computer-based systems' Questionnaire

iii) 'Comparing the generated systems with hand-built systems' Questionnaire

Name ________________________________

Previous computer experience ________________________________
____
____
____
____

The IDDA Generator

Please select ONE answer from each of the following categories

Menus	Very easy to use Easy to use Sometimes easy, sometimes difficult Difficult to use Very difficult
Instructions	Very easy to follow Easy to follow Sometimes easy, sometimes difficult Difficult to follow Very difficult
Diagrams	Very easy to follow Easy to follow Sometimes easy, sometimes difficult Difficult to follow Very difficult
Explanations and error messages	Very easy to understand Easy to understand Sometimes easy, sometimes difficult Difficult to understand Very difficult

Help	Very easy to use Easy to use Sometimes easy, sometimes difficult Difficult to use Very difficult

Help referenced	All the time Frequently Quite often Not much Never

General layout of screens	Too much information presented at one go Sufficient information given Not enough information presented

Were the error messages and the standard messages easy to locate and read?	YES	NO
Did the screens appear cluttered?	YES	NO
Did the screen colours cause confusion?	YES	NO
Did using the second screen cause problems?	YES	NO
Did the responses and actions of the computer appear consistent?	YES	NO

Speed of the computer	Too fast Okay Too slow

Generally found operating the Generator	Very easy Easy Sometimes easy, sometimes difficult Difficult Very difficult

Did you feel confident using the Generator?	Very confident Quite confident Confident Not very confident Not confident

When did you begin to feel happy/confident using the Generator	After the first couple of screens Near the beginning Half way through Near the end Never

Please fill in the following sections (if more space is required, please attach further sheets of paper).

Difficulties encountered when operating the Generator
(Please specify) ______________________________

____
____
____
____
____
____
____
____
____
____
____
____
____
____
____

Suggestions for improvements (Please specify) ____________________

Any further comments (Please specify) ____________________

Time taken to produce the Assessment Questionnaires ____________

Time taken to produce the User-defined Help ____________

Time taken to develop the IDDA end-system ____________
using the Generator
(if more than one session, enter the total time required)

Number of questions in the various assessments

Stage One	________
Stage Two	________
Stage Three	________

Total number of User-defined Help sections linked in ________

Signed ________________________ Date ____/____/____

Benefits of computer-based systems

Will clinical care improve by using such systems?
If Yes, how will it be improved? If No, what kind of system would be required?

Would using such a system interfere with the patient-physician relationship?

What would be the purpose of computer systems in specialist medical fields, as you see it?

Would such a system fit easily into your working practice or would it delay work unnecessarily?

How could the system be improved to assist you further?

Comparing the generated systems with hand-built systems

Does the system built by tools accomplish the required job?

How does the system compare with the hand-built system?

Does the system have extra features that are useful?

Does it lack features that the hand-built system has?

Further comments